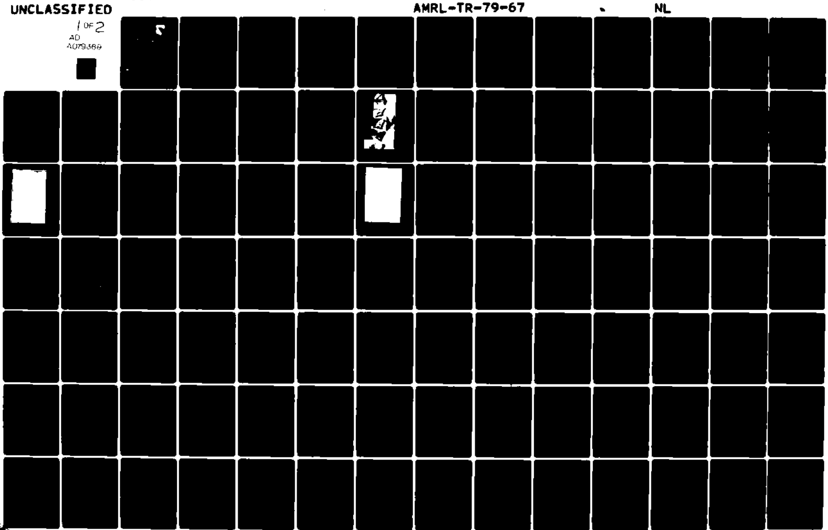
AD-A079 369 SYSTEMS RESEARCH LABS INC DAYTON OH ELECTRONIC WARFA--ETC F/G 1/3
DEVELOPMENT OF A VISUAL INSPECTION TECHNIQUE (OPTICAL ASSESSMEN--ETC(U)
OCT 79 F E WARD , A J DEFRANCES F33615-77-C-0535

UNCLASSIFIED AMRL-TR-79-67 NL

1 OF 2
AD
A079369



LEVEL II

(12)

ADA079369

AMRL-TR-79-67

DEVELOPMENT OF A VISUAL INSPECTION TECHNIQUE (OPTICAL ASSESSMENT OF AIRCRAFT TRANSPARENCIES)

FRANK E. WARD
ANTHONY J. DeFRANCES
ELECTRONIC WARFARE CENTER
OPTRONICS DEPARTMENT
SYSTEMS RESEARCH LABORATORIES, INC.
2800 INDIAN RIPPLE ROAD
DAYTON, OHIO 45440

ROBERT G. EGGLESTON
AIR FORCE AEROSPACE MEDICAL RESEARCH LABORATORY

OCTOBER 1979

Approved for public release; distribution unlimited.

AIR FORCE AEROSPACE MEDICAL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICES

TECHNICAL REVIEW AND APPROVAL

AMRL-TR-79-67

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

CHARLES BATES, JR.
Chief
Human Engineering Division
Air Force Aerospace Medical Research Laboratory

AIR FORCE/56780/11 December 1979 — 150

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
AMRL-TR-79-67

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
DEVELOPMENT OF A VISUAL INSPECTION TECHNIQUE (OPTICAL ASSESSMENT OF AIRCRAFT TRANSPARENCIES).

5. TYPE OF REPORT & PERIOD COVERED
Technical

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Frank E. Ward
Anthony J. DeFrances
Robert G. Eggleston*

8. CONTRACT OR GRANT NUMBER(s)
F33615-77-C-0535

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Electronic Warfare Center/~~Optronics Dept.~~
Systems Research Lab~~oratories~~, Inc.
2800 Indian Ripple Road
~~Dayton, Ohio 45440~~

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61102F, 2313-V1-25

11. CONTROLLING OFFICE NAME AND ADDRESS
*Air Force Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson AFB, Ohio 45433

12. REPORT DATE
Oct[redacted]979

13. NUMBER OF PAGES
129

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Visual inspection
Visual psychophysics
Aircraft windscreens
Optical assessment

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
This work assessed the utility of different types of targets and psychophysical procedures for evaluating optical distortions induced by aircraft windscreens. Targets, both static and dynamic, were viewed through windscreens and the amount of distortion was judged by a magnitude estimation procedure. Judgments were analyzed by discriminant analysis to identify the targets that best facilitated good discrimination among windscreens. A psychophysical matching procedure was also evaluated. In some conditions, photographic representations

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

of the distortion patterns were evaluated using the magnitude estimation procedures. The results of the work show that windscreen-induced distortion is a multidimensional attribute and is best evaluated by multiple inspection procedures. Specific static and dynamic targets are recommended for use in evaluating distortion. Correlations between physical measures of distortion and psychophysical judgments are reported as well as reliabilities for selected experiments. Suggestions for improvements and further work are included.

PREFACE

The research reported in this document was initiated by the Air Force Aerospace Medical Research Laboratory (AFAMRL), Human Engineering Division, Wright-Patterson Air Force Base, Ohio. The experiments were conducted by the Optronics Department, Systems Research Laboratories, Inc., (SRL), Dayton, Ohio, under Air Force Contract F33615-77-C-0535. Dr. Frank E. Ward and Mr. Anthony J. DeFrances were the principal investigators and Mr. Harold L. Iffland was the project manager. The contract was monitored by Major Robert Eggleston of AFAMRL Visual Display Systems Branch.

The authors express their appreciation to the following people who contributed to the success of the project: Mr. Robert McIntyre (SRL, subject procurement and management); Ms. Sharon Ward (SRL, statistical analysis implementation and interpretation); Mr. Clarence Randall and Mr. Jeff Pitsinger (SRL, Photo Lab); Mr. Joe Jordan, Ms. Maureen Deis, and Ms. Carol Brown (SRL, Graphics Department); and Mr. Gary Gomes (SRL, electronic support). Special acknowledgements are extended to Ms. Ursula McCarthy (SRL, research assistant) who scheduled and conducted all data collection sessions; and to Major Robert Eggleston (AMRL/HEA, project monitor) for his cooperation and assistance in providing research facilities.

This technical report was submitted by the authors in March 1979.

A

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (cont.)

LIST OF TABLES

LIST OF TABLES (cont.)

LIST OF TABLES (cont.)

SECTION I
INTRODUCTION

1. GOALS AND SCOPE

This project addressed the problem of assessing the optical quality of aircraft windscreens from a psychophysical perspective. SRL's task was to develop a reliable visual inspection technique that can be applied by quality control inspectors in a manufacturing setting. The task required targets for use in the visual evaluation process be designed, evaluated, and utilized along with appropriate psychophysical procedures in a windscreen evaluation process.

2. GENERAL

Section I of this report reviews the background of optical quality assessment by psychophysical methods. It describes SRL's approach to the problem and explains the rationale for a solution.

Section II presents the data from the Phase I pilot work. Phase I was concerned with evaluating the suitability of a variety of targets for use in the visual inspection process.

Section III focuses on the larger Phase II studies. This work was concerned with in-depth testing of targets selected from the Phase II studies. In addition, alternative psychophysical methods were tested.

Section IV reports the results of a series of studies designed to assess the feasibility of rating windscreen distortion by the use of photographs taken through the windscreens. Ratings on photographs were compared with judgments made on actual windscreens.

Section V reports the results of a study that utilized a matching technique to rate distortion. The work focused on relating windscreen distortion to physical measures.

Section VI contains a general summary, conclusions, and recommendations.

3. BACKGROUND

The development of visual inspection techniques to assess the optical quality of aircraft windscreens is a complicated problem because of the nature of the windscreens. Windscreens for high performance aircraft are generally curved, frequently thick, conform to a complex geometry, and may contain several laminated layers. Extremely small variations in the radius of curvature of the inner and outer surfaces produces a variety of optical defects which are exaggerated in actual viewing because the windscreens are installed at a shallow angle of incidence with respect to the observer's line of sight.

Since the manufacturing process is not perfect, some amount of optical distortion will probably always be present in curved windscreens designed for high performance aircraft. The manufacturer and, of course, the Air Force must be able to judge the severity of optical distortion in order to determine the quality and suitability of each windscreen for operational use.

Ideally, distortion would be quantified by physical measurement. There are objective measures of optical quality such as line-slope deviation, lens factor, and displacement grade. The difficulty is that when these objective measures are used to define the optical acceptability of a windscreen, actual visual inspection might suggest that a particular windscreen is unacceptable even though it meets the physical standards or vice versa. That is, the existing objective physical measurements do not always reflect the subjective visual perception that an observer experiences. These discrepancies probably occur because the physical measurements that have been developed only provide an index of the maximum magnitude of distortion found in the windscreen. In a strict sense, this value only describes distortion in the small area of the windscreen where the measurement is taken. Unfortunately, all of the physical measurement techniques are insensitive to changes in the severity of distortion throughout the entire windscreen area. Yet these changes produce a pattern of distortion which is clearly visible and which no doubt influences one's visual impression of distortion. Until the optical engineers can develop measurement techniques to capture the total distortion effect in windscreens of complex geometrical design, we will have to rely on using visual inspection for assessing optical quality. Therefore, it is important that the visual inspection be carried out in a standardized and systematic manner.

To date, there is no industry-wide standardized procedure for visual inspection. Grid patterns are commonly used but it is not known if these targets are the best ones to highlight optical distortion. In addition, visual inspectors are generally free to use whatever procedure they desire to inspect the windscreens. Naturally, this has resulted in a variety of idiosyncratic inspection methods. While some may be valid, little data exist to suggest what procedure is best. Thus we are faced with two problems: how should visual inspection be conducted; and, what targets best highlight optical distortion?

Previous work by Gomer and Eggleston (1978) has shown that direct magnitude scaling can be successfully applied to the windscreen inspection problem. These investigators had subjects rate windscreens in terms of optical distortion and found the correlation with the objective measures, lens factor and displacement grade, to be surprisingly high (although some differences were noted). In their study, the subjects' ratings tended to agree. However, Gomer and Eggleston did not have sufficient time to accurately determine the reliability of the magnitude estimation procedure. Further, they were restricted to using an available grid target, which was designed for use in photographing windscreens. This target, white lines on a black background, while suitable to highlight distortion on a photograph, may not have been optimal for capturing the subjective global effects of distortion perceived by a quality control inspector or air crew personnel.

4. APPROACH

The approach to the problem was to conduct a series of laboratory studies. Selection of potential targets to highlight distortion was based on the subjective judgments of the investigators. With two exceptions, theoretical means were not used to select targets because no suitable mathematical analysis existed which could capture the global effects of distortion. This is the very reason for the present study.

With the aid of the AMRL windscreen laboratory personnel, the investigators selected five windscreens that seemed to be representative in exhibiting the range of optical distortions commonly encountered in the manufacturing

process. The choice of windscreens was restricted, however, by the availability of samples from the windscreen laboratory.

The experimental work was divided into two phases. Phase I was devoted to pilot work to select potential targets to be used in the inspection process. Most experiments were carried out using only three or four subjects. The suitability of test targets for Phase II was determined by several considerations. The targets that seemed to best allow discrimination of the severity of distortion among the five windscreens furnished by AMRL/HEA were chosen for further study. In addition, some targets and target conditions were selected because those particular test conditions seemed subjectively to facilitate judgments of distortion.

In Phase II, the best targets were again evaluated, however, a larger number of subjects (10) were used in the inspection process. Additional investigations were also carried out to ascertain the difference, if any, between real-time assessment of distortion and judgments made solely on the basis of a photographic record. A correlational analysis was performed to determine the relationship of physical measures of distortion with those obtained by psychophysical techniques. In addition, an analysis of the reliability of different techniques was performed.

Most of the experiments reported in this document were concerned with determining target suitability. Suitable targets are those which enhance optical aberrations so that observers can reliably discriminate among the optical distortions found in representative windscreens. Once targets and appropriate psychophysical procedures have been developed, they can be combined to produce a visual inspection technique suitable for use during quality assessment of aircraft windscreens.

SECTION II

PHASE I EXPERIMENTS

1. SUBJECTS

The subjects used in all of the studies were recruited from the AMRL paid subjects pool. They were naive with respect to the purpose of the windscreen studies but were experienced in work involving visual-tracking displays. Subjects were given optometric eye examinations to insure that they met the vision standards for Category II flying status.

2. APPARATUS

Except where noted, the basic apparatus was the same for all studies reported. Five Government-furnished F-111 windscreens were mounted at an installation angle of 21.6 degrees (with respect to the horizontal line of sight) so that the subject could view the targets from the F-111's pilot design eye position. Windscreens were supported by a rigid frame which could be rolled laterally on tracks so that each windscreen could be accurately positioned between the subject and the screen upon which the targets were projected. Each subject was seated in an observation booth and used a chin rest to fix the head position. Masks and templates were attached to the windscreens so that a subject's view included only those sections of the windscreen relevant for study. Figure 1 shows the windscreens mounted on the movable frame in front of the subject's observation booth.

Targets were projected onto an 8-foot square screen by a lantern slide projector that was located behind and above the subject's viewing position. This arrangement minimized projection parallax while giving the subject a reasonably bright image that filled 22.5° of the approximately 32° x 43° windscreen aperture. The relative positions and dimensions of the experimental setup are detailed in Figure 2.

When properly adjusted, the projector filled the screen and produced a center field luminance of 5 footlamberts (FL). Falloff towards the edge of the screen never exceeded 17% of the center field luminance value. The light areas of the target negatives had an average density of 0.34, hence effective luminance was 2.3 FL. The dark areas of the targets averaged 0.76 FL,

Figure 1. The Movable Windscreen Support Stand and Subject's Observation Booth

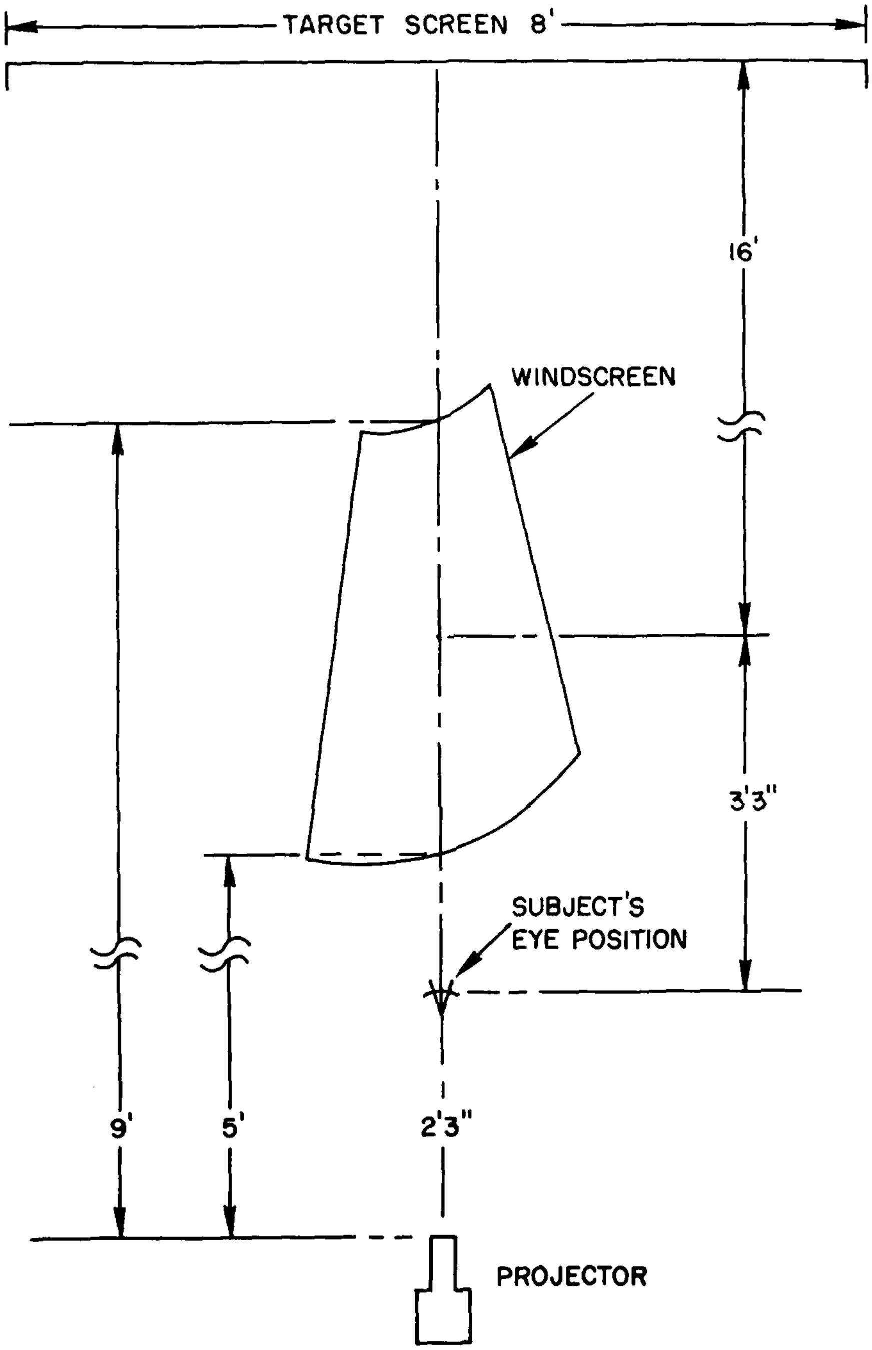


Figure 2. Overhead View of the Experimental Layout

producing a contrast of approximately 50%. The windscreens varied in their light attenuating properties; the average windscreen density was about 0.28 at the boresight.

3. PROCEDURE

The subject's task was to judge the symmetry or regularity of the target pattern viewed through the windscreen. For comparison, subjects were told to assume that a target viewed without a windscreen in place would be perfectly symmetric or undistorted. Observers were instructed to report the fraction or percentage of the test pattern that was symmetric and were permitted to use a rating scale of their choice, such as decimals, fractions, or percentages*. They were further told not to try to count or quantify undistorted target elements, but rather to give their first, global impression. For practice, observers were shown a target pattern photographed through a windscreen and were asked to make a judgment. Questions were answered and, after they appeared to understand the task, subjects were seated in the observation booth and testing was begun.

During changes of targets and/or windscreens, the subject's view was occluded. Windscreens were randomly presented across blocks of replicates for each target. Subjects always judged the same five windscreens five times each for a given target. Thus, there were 25 judgments made per subject for each target.

Appendix A contains instructions to the subjects for the different categories of experiments.

4. ANALYSIS

Data were keypunched and a stepwise discriminate analysis was performed. This analytical technique was chosen because not only is it suitable for

---

*Some subjects preferred to judge the proportion of nonsymmetry or target distortion. In principle, this approach seemed the equivalent (though the inverse) of the one suggested, so it was permitted. The proportion judgments of nonsymmetry were transformed by subtracting from 1.0 so that all proportions had large values associated with minimal target distortion.

analysis of single-target data but it is also especially useful for multiple-target analyses. The procedure finds the best linear combination of variables (targets) that allows optimal differentiation among groups (windscreens).

The analytical program, BMD07M*, performs a multiple group discriminate analysis, then computes a set of linear classification functions in a (user-specified) stepwise fashion. The procedure used by the program is essentially an extension of multivariate analysis of variance and stepwise regression. At each step, a multivariate analysis is used to determine if there is a significant difference among the groups (windscreens) for the variables (targets) under consideration. Then, if a significant difference exists, the program determines the linear combination of variables that produces the best discrimination among the groups.

If only one variable (target) is evaluated, its utility for discrimination is assessed by an F-statistic for differences among group means. This is the same procedure as a one-way repeated measures analysis of variance. Discrimination among windscreens is given by the ranking and differences among group means.

When two or more variables (targets) are simultaneously evaluated, the program selects variables for discrimination in a stepwise fashion. There are several criteria that can be used for selection of the variable to be included or deleted at each step. The most commonly used procedure employed the F-statistic for differences among group means. The first variable chosen was the one which had the highest F-statistic. At each subsequent step, an F-statistic is computed for each variable not already included in the discriminant function. Each F-statistic corresponding to a particular variable reflects the amount of discrimination afforded were that variable included in the discriminant function. At each successive step the variable with the largest F is added to the analysis.

A second procedure to select variables for the discriminant function ignores the F-statistic and includes variables that best separate the most

---

*See Dixon (1977) for a description of the program.

similar groups. This procedure was used only for in-depth analysis, after an analysis using the F-statistic criterion had been performed.

The output of the discriminant analysis program affords several alternative procedures for interpreting the data. First, a classification chart is produced. This chart shows actual group membership versus group membership predicted from measurements obtained from the target(s) under consideration. Table 2 of Experiment I (on page 22) is an example of this classification. The diagonal entries are correct classifications. Off-diagonal elements show how observations were incorrectly classified. A simple index of the discrimination quality of a variable or group of variables (targets) can be computed by taking the ratio of correct classification to the total sample size. Thus, for Table 2, 22 correct classifications out of 30 possible gives $(22 \div 30) \times 100 = 73.3\%$ correct classification. This procedure is only approximate because a particular observation may have very similar probabilities for membership in several different groups. It will be classed in the group with the highest probability, and no indication will appear in the table that this observation could easily have been classified differently. This situation is analogous to saying that 0.499 belongs to a different group than 0.501 because 0.499 has a value less than a cutoff point of 0.500. Clearly, the two values are more similar to each other than, say, 0.250 and 0.750. Thus, the true discriminator may be somewhat better or worse than that actually indicated.

A second way to interpret the data is to examine the canonical variates. A canonical variate is a linear combination of dependent variables (judgments of windscreen quality). The first canonical variate is the linear combination of targets that best discriminates among groups (windscreens). Each successive canonical variate is orthogonal to all the preceding ones, and discriminates among windscreens as well as possible under that restriction.

The number of canonical variates that can be extracted depends on the number of targets. The maximum number is the number of targets used for discrimination. In practice though, there is usually little to be gained by examining more than two canonical variates. Often the first two variables

together account for more than 99% of the dispersion in the judgments selected for entry into the discriminant function.

When the first canonical variate accounts for a large percentage of the total dispersion (variability), whatever relationship the judgments and targets have is largely explained. Dispersion is analogous to the "proportion of variance accounted for" in the univariate analysis of variance*. As additional canonical variates are considered, they will tend to have lower eigenvalues, and hence, lower canonical correlations. This is an indication that the values of the dependent variable (judgments) are correlated with each other. This is just another way of saying that the pattern of judgments is redundant and provides little additional information about windscreen discrimination. Formally, dispersion for a given canonical variate is expressed as the ratio of the eigenvalue for that variate to the sum of the eigenvalues of the analysis.

A third way to examine the data is afforded by the canonical correlations. For each canonical variate, the canonical correlation gives the relationship between the dependent variables (judgments) and the independent variables (windscreens). In the case of a single variate, the square of the canonical correlation represents the proportion of dispersion (variability) that the single target accounts for. When there are several canonical variates, each variate's canonical correlation indicates how well the judgments and windscreens are related for that variate. In general, the canonical correlations diminish as further variates are extracted, reflecting the fact that each successive variate is less likely to contribute to discrimination.

Discriminant analysis is a generic term for any multivariate analysis that attempts to group data by some rule and, as such, is a broad topic. A variety of techniques can be used to establish rules for discrimination; the program used, BMD07M, is but one of many discriminant analyses. The reader who wishes to pursue the topic in detail should refer to Marriott (1974) for

*In this context it is important to note that even though a canonical variate may be statistically significant in accounting for differences among group means, the variate may still only account for a small amount of the total variability. Statistical significance should not be confused with the variability accounted for by a given linear combination of targets.

an overview. Rao (1965) presents the detailed mathematical background for an in-depth understanding of the linear algebra involved in discriminant analysis.

The three techniques mentioned above all provide the experimenter with different perspectives with which data can be examined. In the analyses of experiments that follow, not all of the analytical methods are cited for each study. Only those deemed most informative are reported.

5. RESULTS

The discriminate analysis program described above requires data whose errors are approximately normally distributed. Preliminary analysis showed that this could be achieved by using an arcsin square root transformation. The transformed data were then Z-scored to enable intersubject comparisons to be made. Some representative data on these distributions are given in Appendix B.

The data for each experiment are plotted and the results of the discriminate analysis are shown. Plots show the mean Z-scope for each windscreen with positive scores indicating the least perceived windscreen distortion. The discriminate analysis classification chart shows, for each windscreen, the number of times it was classed correctly and the number of times it was incorrectly classed as one of the other windscreens. When appropriate, plots of the first two canonical variates are shown and the canonical correlations are indicated.

6. EXPERIMENT I: LINE GRIDS

For this first experiment only, windscreens were supported by a frame that provided two viewing apertures. One aperture always contained a windscreen and the other was used without a windscreen for the nondistorted control. Subjects were instructed to view the target through the open aperture, then to look through the test windscreen and make judgments. The test windscreen was changed for each trial. This procedure proved to be very cumbersome, so a five-aperture movable support stand was constructed and used for all further work. It was found that subjects were able to make judgments of distortion without the necessity of repeated inspection of the nondistorted

target. All but two targets contained periodic or rectilinear elements that seemed to structure the perception of symmetry. Even though part of the visual array was distorted, there was a "perceptual expectation" of symmetry against which subjects could make their judgments.

In the first experiment, six subjects judged windscreen distortion while viewing four different grid size targets in positive and negative contrast configurations: grids with white lines on a black background and black lines on a white background were used. Figure 3 shows an example of a positive contrast grid target. Grid sizes, in degrees of visual angle, are shown in Table 1.

TABLE 1. LINE GRID SIZES IN DEGREES OF VISUAL ANGLE

Grid Number	Lines/Color	Background Squares/Color
1	.015 Black	.080 White
2	.035 Black	.180 White
3	.040 Black	.275 White
4	.060 Black	.370 White
5	.015 White	.080 Black
6	.035 White	.180 Black
7	.040 White	.275 Black
8	.060 White	.370 Black

According to discriminant analysis, the best discrimination among the five windscreens was produced when grid number 3 was the visual target. The windscreen classification chart is shown in Table 2. According to the analysis, 73% of the windscreens were correctly classified.

TABLE 2. CLASSIFICATION CHART FOR JUDGMENTS BASED ON GRID 3

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	2	3	0	1	0
B	3	3	0	0	0
C	0	0	6	0	0
D	0	0	0	6	0
E	0	0	0	1	5

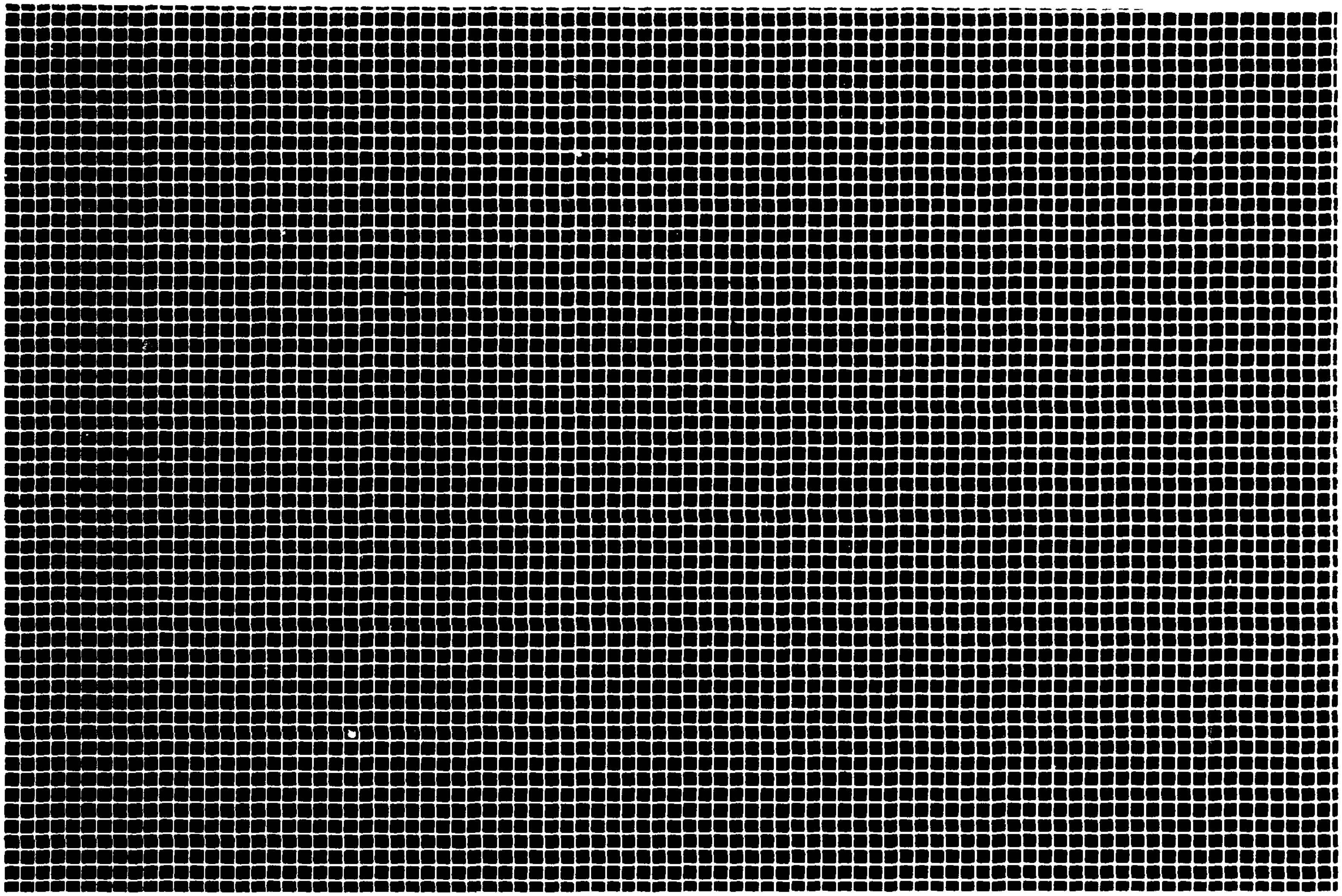


Figure 3. An Example of a Positive Contrast Grid Target

The mean Z-scores obtained for grid 3 are shown in Figure 4. It is apparent that windscreen C was judged to be much poorer than the others. Windscreen B was judged to have the least distortion.

The analytically derived linear combination of two variables produced slightly better discrimination as is shown in Table 3. This analysis combined judgments with the first target type, grid 3, with a second target, grid 7, to produce 80% discrimination.

TABLE 3. CLASSIFICATION CHART FOR JUDGMENTS BASED ON GRID 3 AND GRID 7

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	3	2	0	1	0
B	2	4	0	0	0
C	0	0	6	0	0
D	0	0	0	6	0
E	0	0	0	1	5

Figure 5 shows the Z-score means using grid 7.

Figure 6 is a plot of the two canonical variates produced by the discriminant analysis. These two variates accounted for 99.7% of the total variability in the data for the eight grid types. The canonical correlations for the first and second canonical variates were 0.976 and 0.576 respectively. These figures indicate that the first canonical variate will produce the best discrimination. The second variate accounts for some discrimination, but also reflects other dimensions of the subjects' judgments. The plot clearly shows that the first variate separates windscreens C, E, and D from the others. Windscreens A and B are not separated well on dimension 1 but are spread apart a little on the second dimension. Thus, while classification only improved from 73% to 80%, the addition of a second canonical variable reveals that the factor that aids discrimination among windscreens C, E, and D is different from the one that distinguishes windscreens B from A.

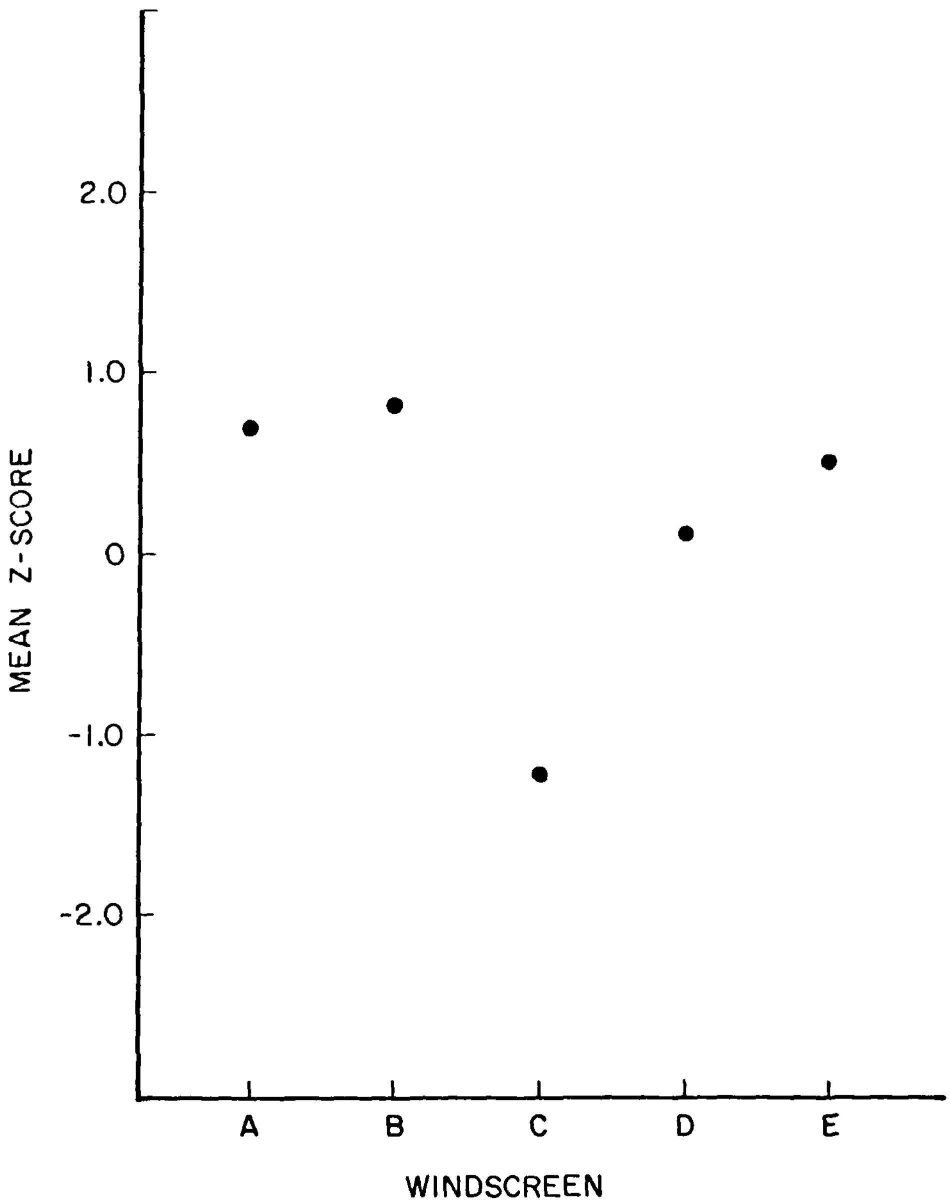


Figure 4. Z-Score Means (Judgments) with Grid 3 as the Target

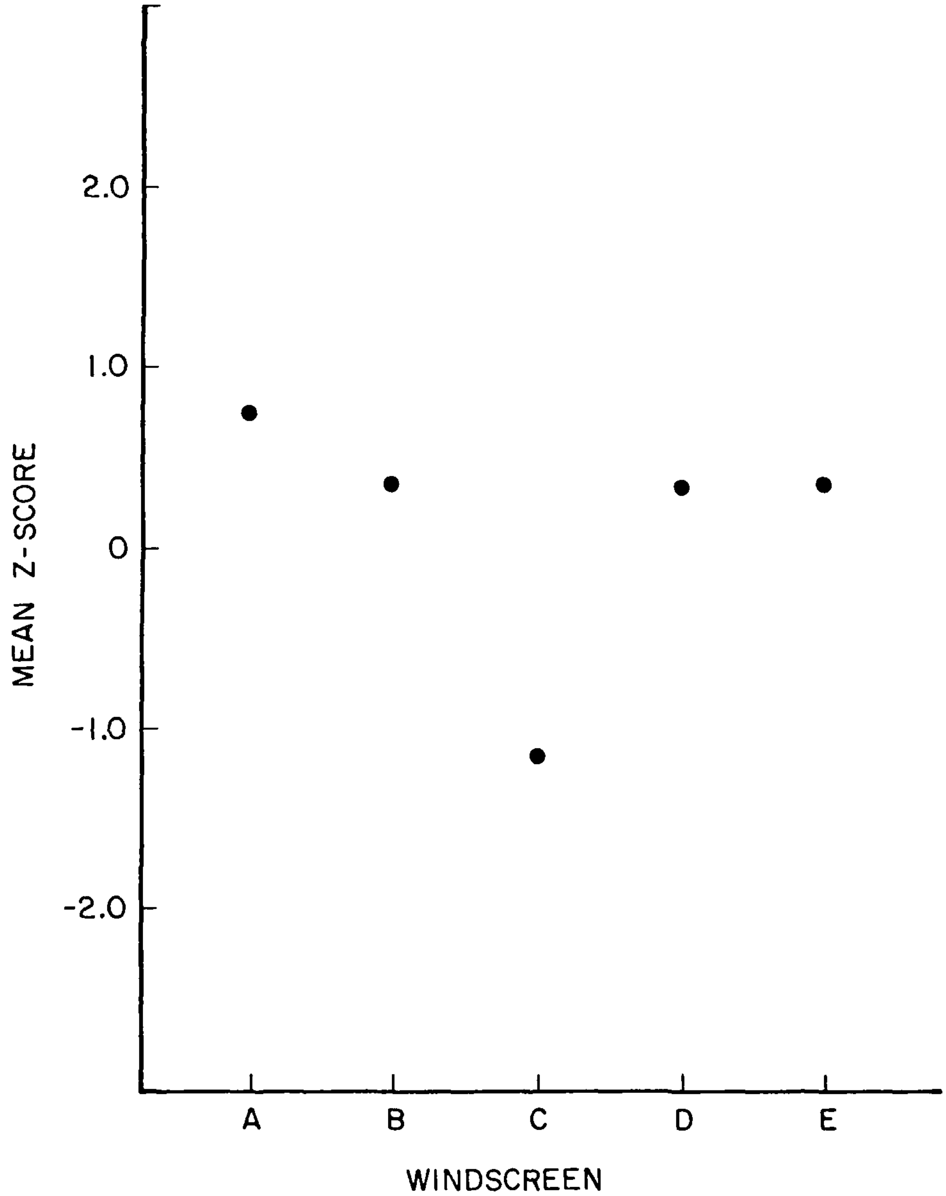


Figure 5. Z-Score Means (Judgments) with Grid 7 as the Target

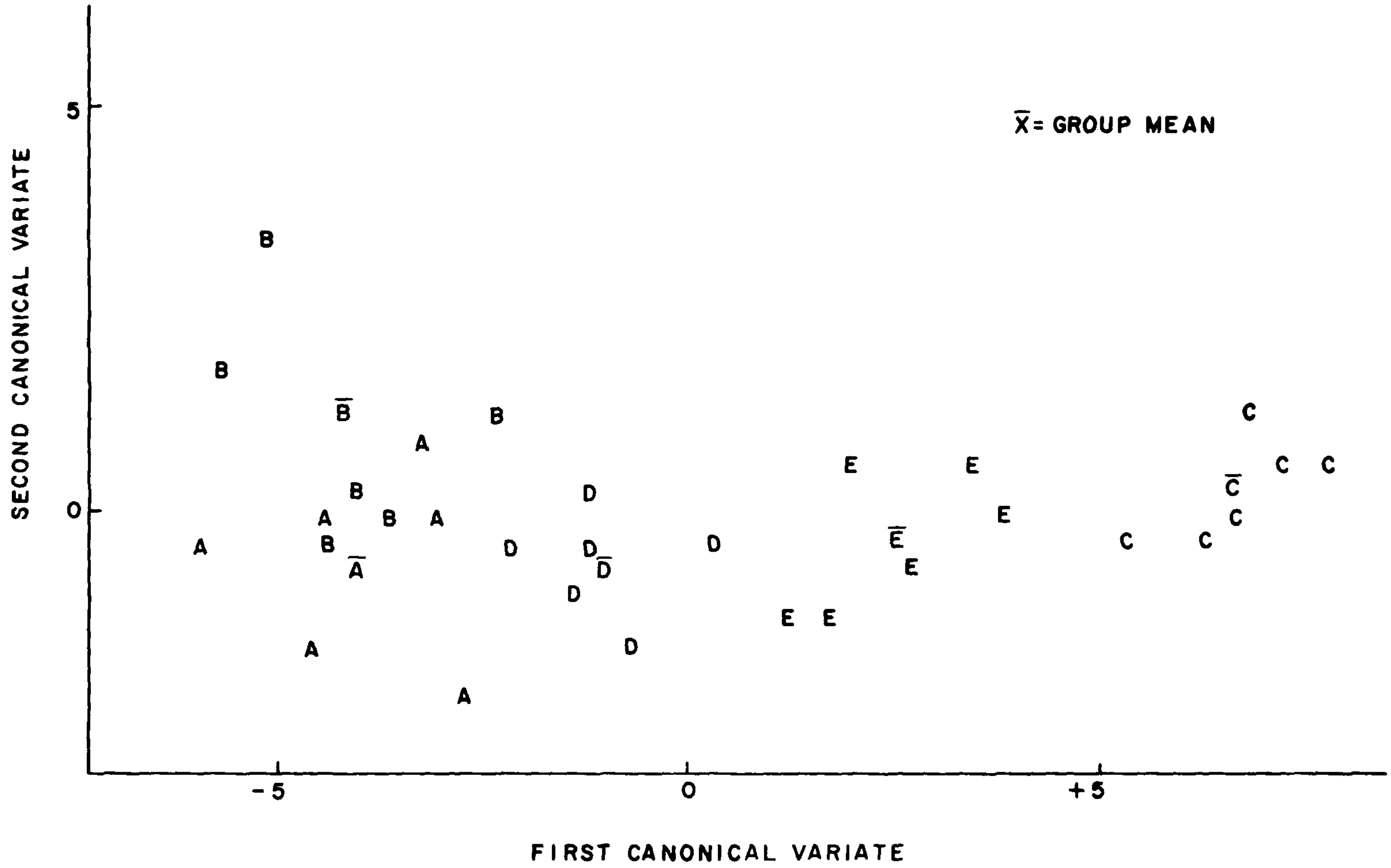


Figure 6. Plot of the First Two Canonical Variates for Experiment I

7. EXPERIMENT II: DOT PATTERN

For this study, three subjects judged a periodic dot lattice pattern. The target consisted of 0.08° black dots on a white background. Dots were spaced about 0.18 degrees visual angle apart and had a contrast* of 0.37. This pattern was chosen because it contained regular, periodic elements, yet avoided the line element patterning of the grids. This target is shown in Figure 7.

Since only one target type was used, only a single canonical variable could be extracted from the data. As is shown in Table 4, 53.3% correct classification was obtained. Windscreens B, D, and E were sometimes classed with windscreen A.

TABLE 4. CLASSIFICATION CHART FOR JUDGMENTS BASED ON THE PERIODIC DOT PATTERN

Windscreen Tested	Windscreen Classed As: A	B	C	D	E
A	0	2	0	0	1
B	1	2	0	0	0
C	0	0	3	0	0
D	1	0	0	2	0
E	1	0	0	1	1

Figure 8 shows the Z-score means for the periodic dot target. Again, windscreen C showed the most distortion. The "best" windscreen was B, although this particular target did little to enhance distortion differences among windscreens A, B, and D.

8. EXPERIMENT III: RANDOM TEXTURE

In this experiment, three subjects viewed four different random texture targets through the windscreens. They judged the proportion of the windscreen area that was distortion-free. These targets were chosen because it was thought that windscreen-induced distortions might cause contrast reductions

*Contrast referred to in this work is defined by

$$C = (L_{max} - L_{min})/(L_{max} + L_{min})$$

where L equals photometric luminance.

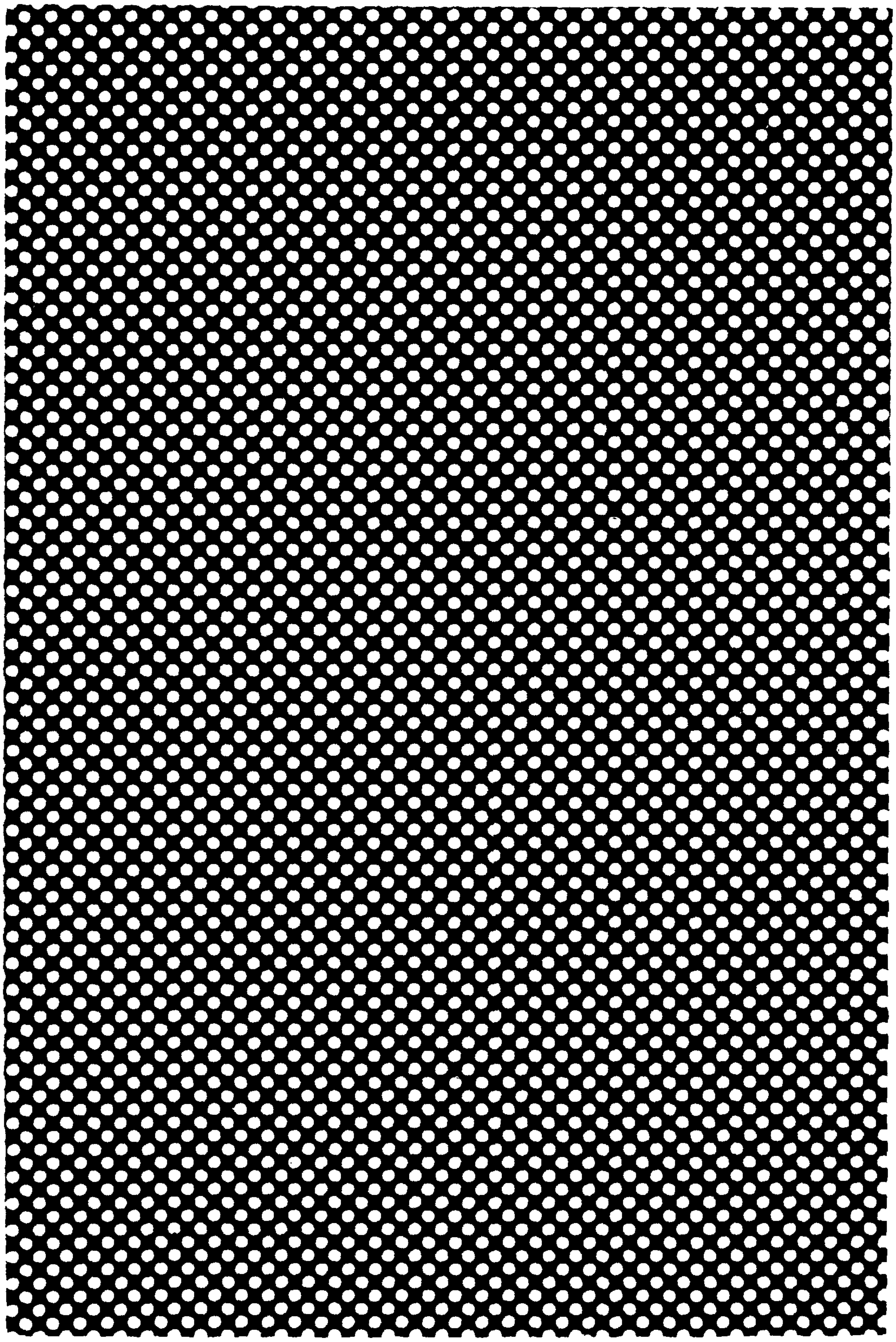


Figure 7. The Periodic Dot Lattice Target

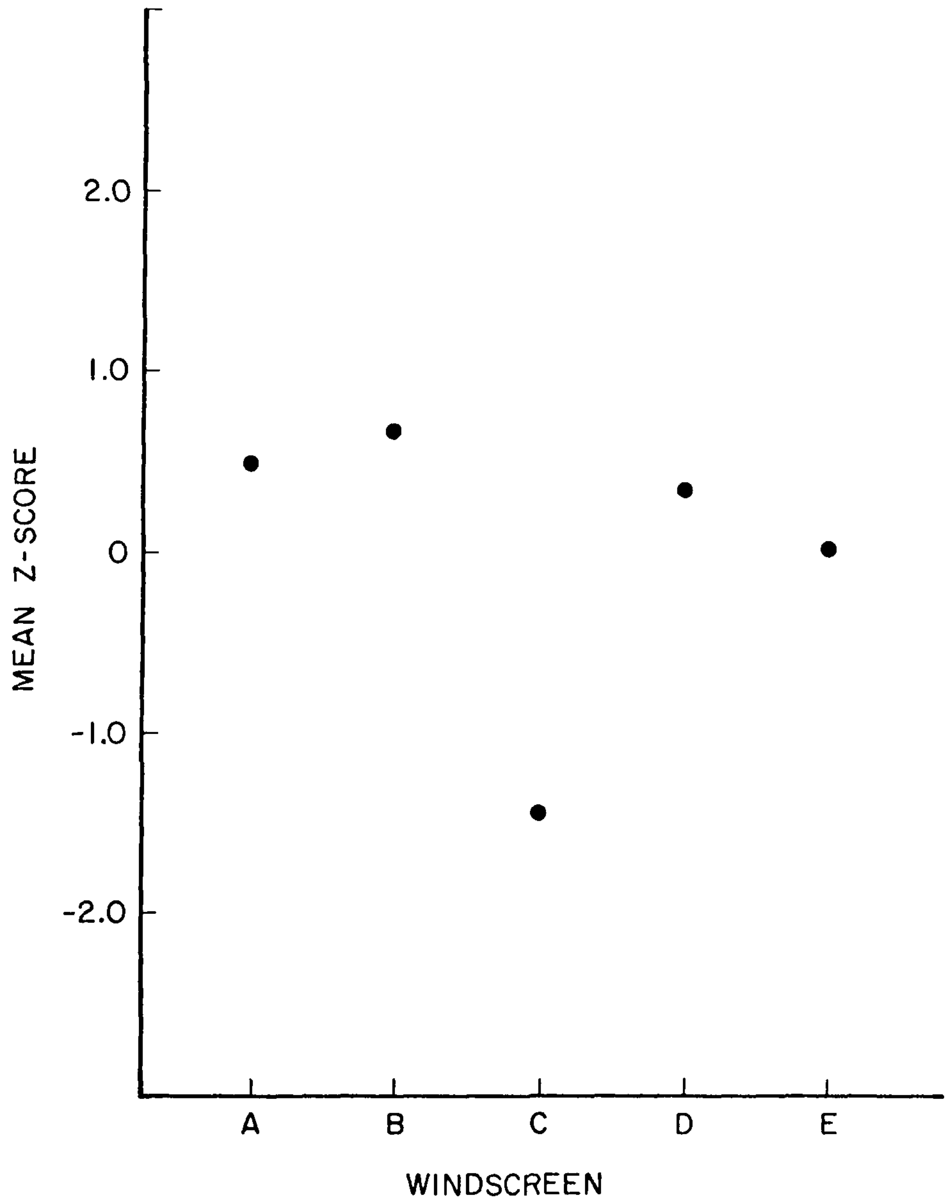


Figure 8. Z-Score Means (Judgments) with the Periodic Dot Target

and/or position alterations such that distorted parts of the target would be below visual resolution. Even though the predicted effect of the distorting windscreen was not dramatic, subjects were able to make meaningful judgments.

It is difficult to describe the targets quantitatively. Most were of medium contrast with different densities of texture elements. Only one target afforded a significant discrimination among windscreens. This target (see Figure 9) was a mezzotint (average contrast 0.76) that averaged 6-7 elements/degree. Element sizes ranged from a maximum of about 0.07° down to about 0.002° in irregular shapes.

The results of this study are shown in Table 5. The classification chart shows that subjects achieved 53.3% correct classification with this target.

TABLE 5. CLASSIFICATION CHART FOR JUDGMENTS BASED ON THE RANDOM TEXTURE TARGET

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	1	0	0	1	1
B	0	2	0	0	1
C	0	0	3	0	0
D	1	0	0	1	1
E	0	1	0	1	1

Figure 10 shows the mean Z-score judgments using random texture target. Windscreen D was judged best when this target was used, although the differences among D, A, and E were not very large. It is apparent that discrimination was about the same for the random texture as for the periodic dot pattern.

These last two targets, the periodic dot and random texture, were selected from several dozen candidates produced by the SRL photographic lab. The authors viewed each target through each windscreen to determine if the target appeared to undergo transformations that captured distortion. Some patterns were rejected because they themselves were found to interfere with a clear perception of the windscreen-induced distortion. Other patterns were discarded either because they seemed inappropriate for assessing the global

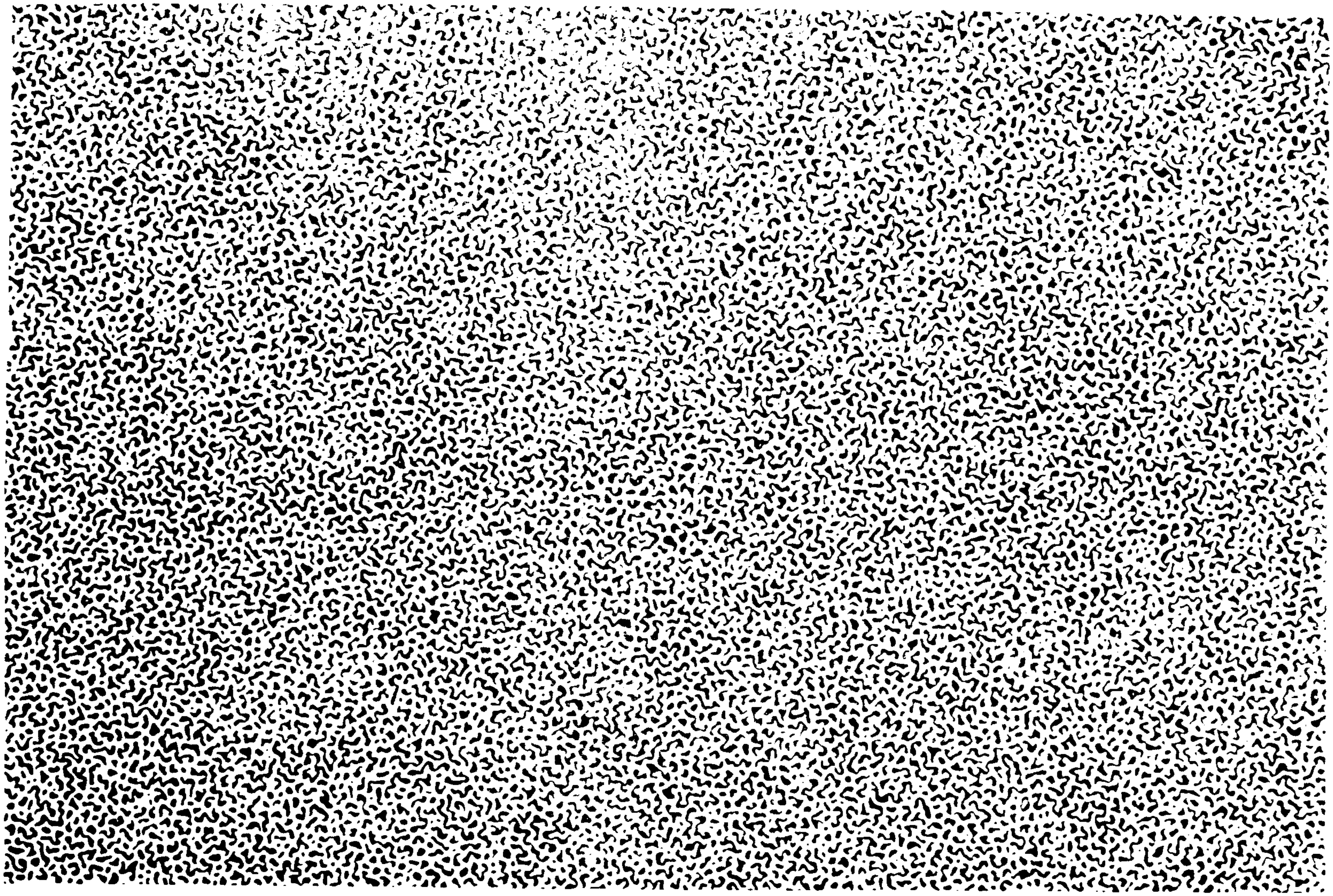


Figure 9. Random Texture Mezzotint Target

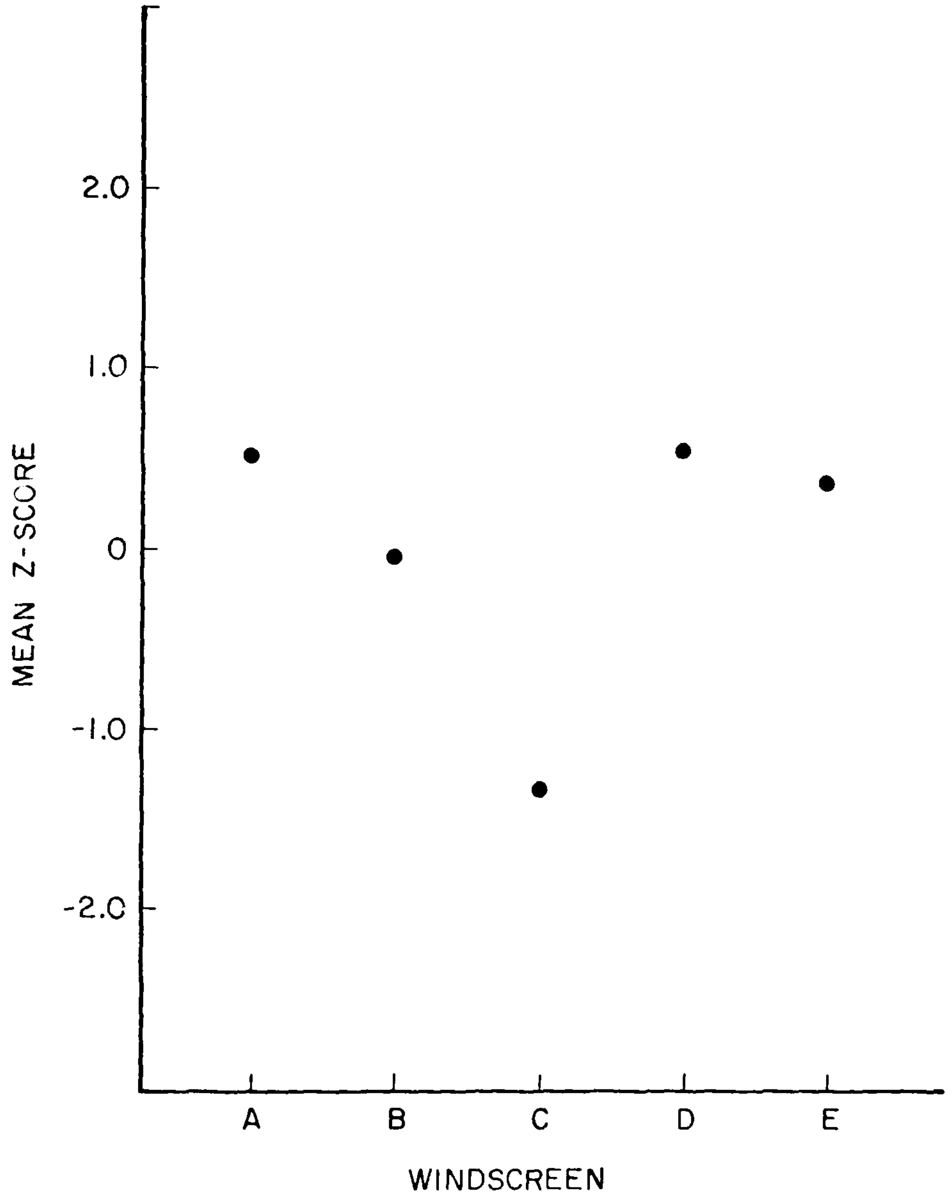


Figure 10. Z-Score Means (Judgments) with the Random Texture Target

effects of distortion or because they did not undergo sufficient deformations when viewed through the windscreens. The targets screened are listed in Table 6.

TABLE 6. TARGETS SCREENED AND FOUND UNSUITABLE FOR ASSESSING WINDSCREEN DISTORTION

Target Type	Number of Targets	Reason for Rejection A	B	C
Seaman Star	2	x		
Moire Patterns	6	x	x	x
Concentric Circles	2			x
Concentric Squares	2			x
Honeycombs	2	x		
Irregular Line Gratings	3		x	
Counterphase Flicker Check	4	x		
Counterphase Flicker Grating	4	x		
Periodic Dot	12			x
Random Texture	12			x
Line Tints	6			x
Herringbone Texture	2	x		
Sinusoidal Displacement Check	4	x		
Sinusoidal Displacement Grid	4	x		

Reasons for Rejection
A: Unusual perceptual effects
B: Inappropriate for global assessment
C: Ineffective

9. EXPERIMENT IV: MERIDIONAL TARGET

This target was one of the targets that was chosen on the basis of a mathematical analysis. The meridional target was designed so that it would exactly compensate for the normal distortion expected to be produced by a conic section such as is found in the F-111 windscreens used in our particular study. The target consisted of lines approximately 0.24° thick arranged as shown in Figure 11. As all the other targets, it was projected so that it filled the 8 x 8-foot screen.

The target did not lead to very good discrimination. As is shown in Table 7, only 46.7% of the windscreens were correctly classified.

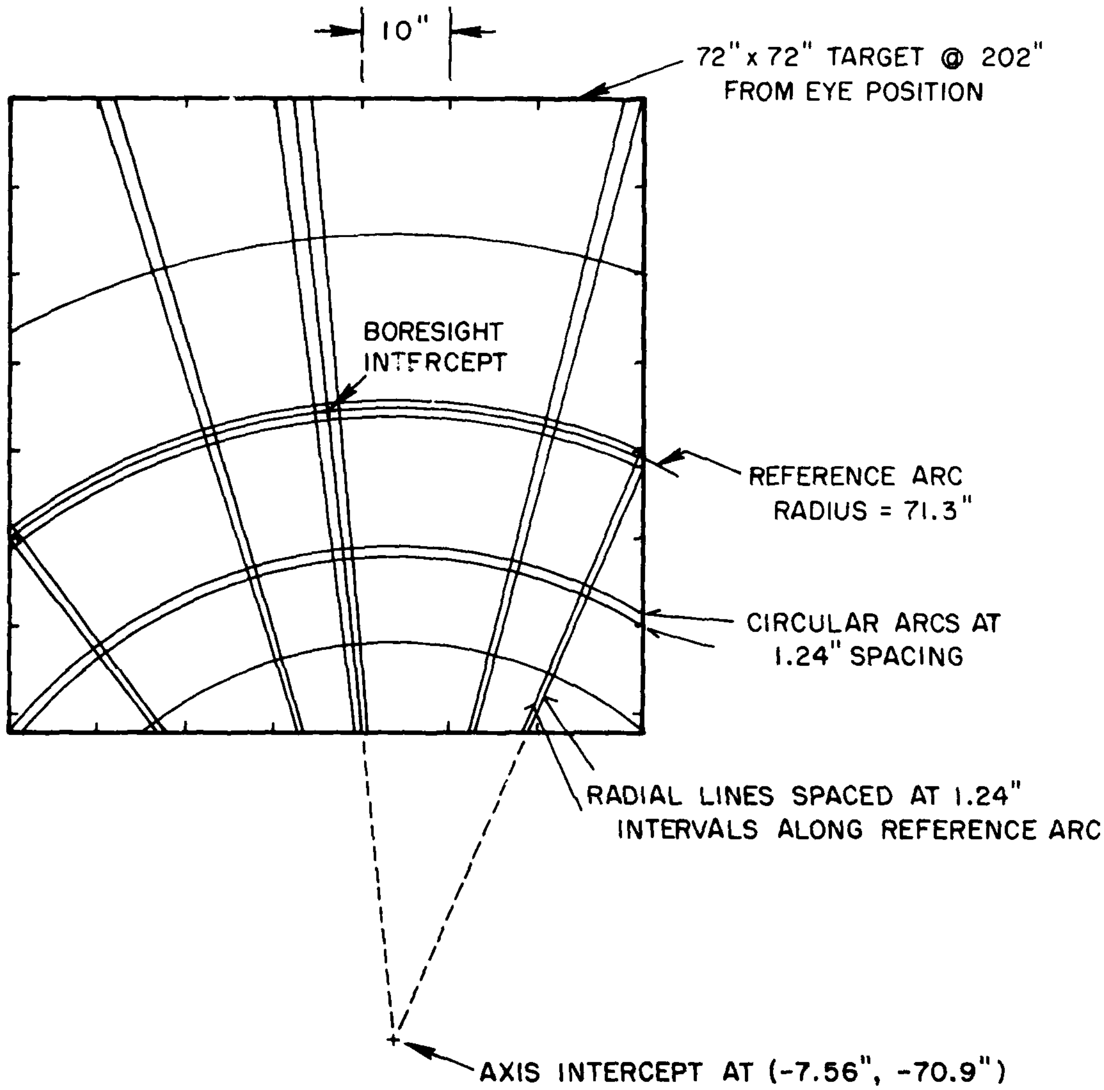


Figure 11. The Meridional Target

TABLE 7. CLASSIFICATION CHART FOR JUDGMENTS BASED ON THE MERIDIONAL TARGET

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	1	0	0	1	1
B	1	2	0	0	0
C	1	0	2	0	0
D	0	1	0	1	1
E	1	0	0	1	1

Figure 12 shows the Z-score means for the meridional target. As in some of the other studies, windscreens A, D, and E are easily confused.

10. EXPERIMENT V: CHECKERBOARDS

In this study, three subjects rated windscreen distortion for four different sizes of checkerboard targets. The target dimensions were chosen to parallel those used in Experiment I, the grid study. The checkerboard dimensions are given in Table 8 and a typical example is shown in Figure 13.

TABLE 8. DIMENSIONS FOR THE CHECKERBOARD TARGETS

Check	Degrees Visual Angle of White/Black Squares
1	0.08
2	0.19
3	0.28
4	0.37

Table 9 shows the classification achieved for the "best" checkerboard target, check 4. This target yielded a correct classification of 73.3%.

TABLE 9. CLASSIFICATION CHART FOR JUDGMENTS BASED ON CHECK 4

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	2	1	0	0	0
B	1	2	0	0	0
C	0	0	3	0	0
D	0	0	0	2	1
E	0	0	0	1	2

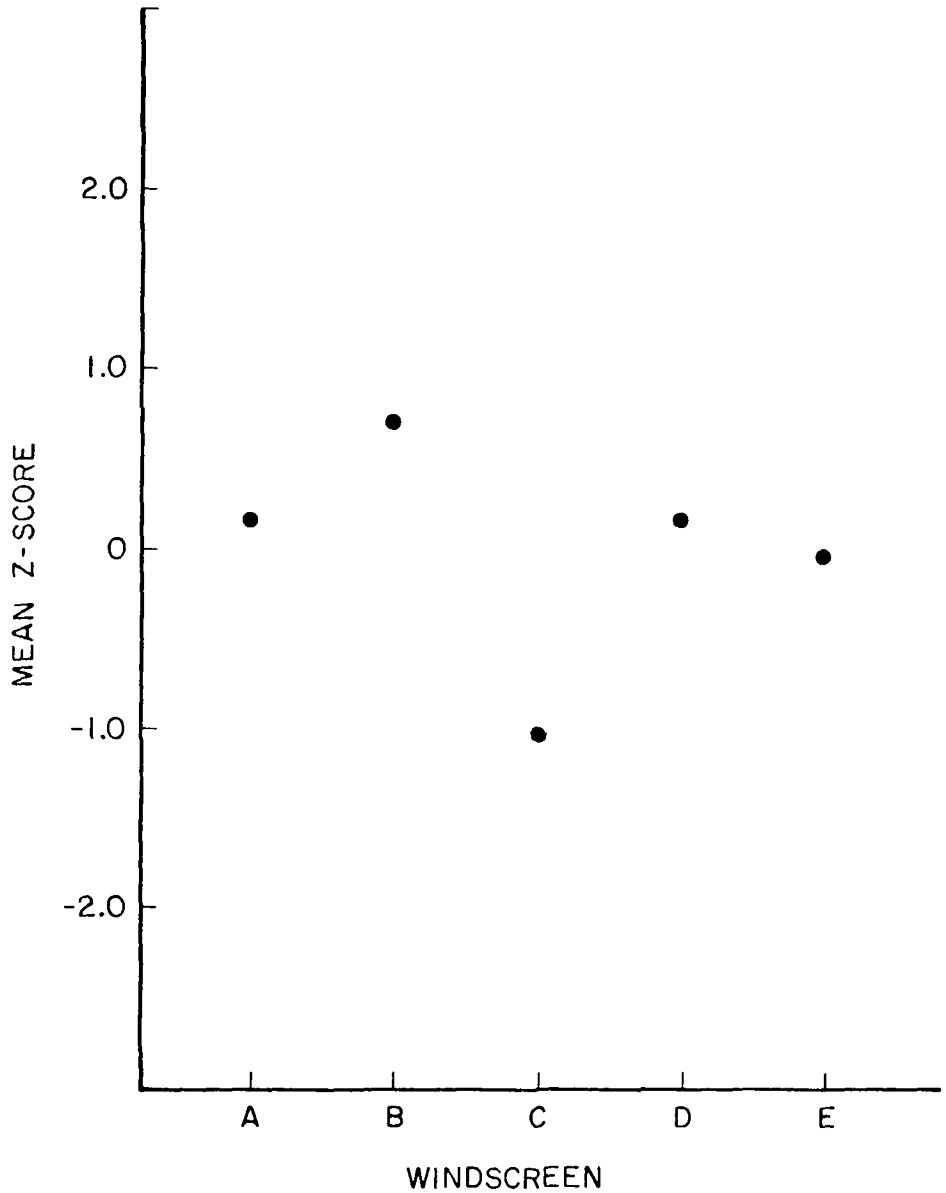


Figure 12. Z-Score Means (Judgments) with the Meridional Target

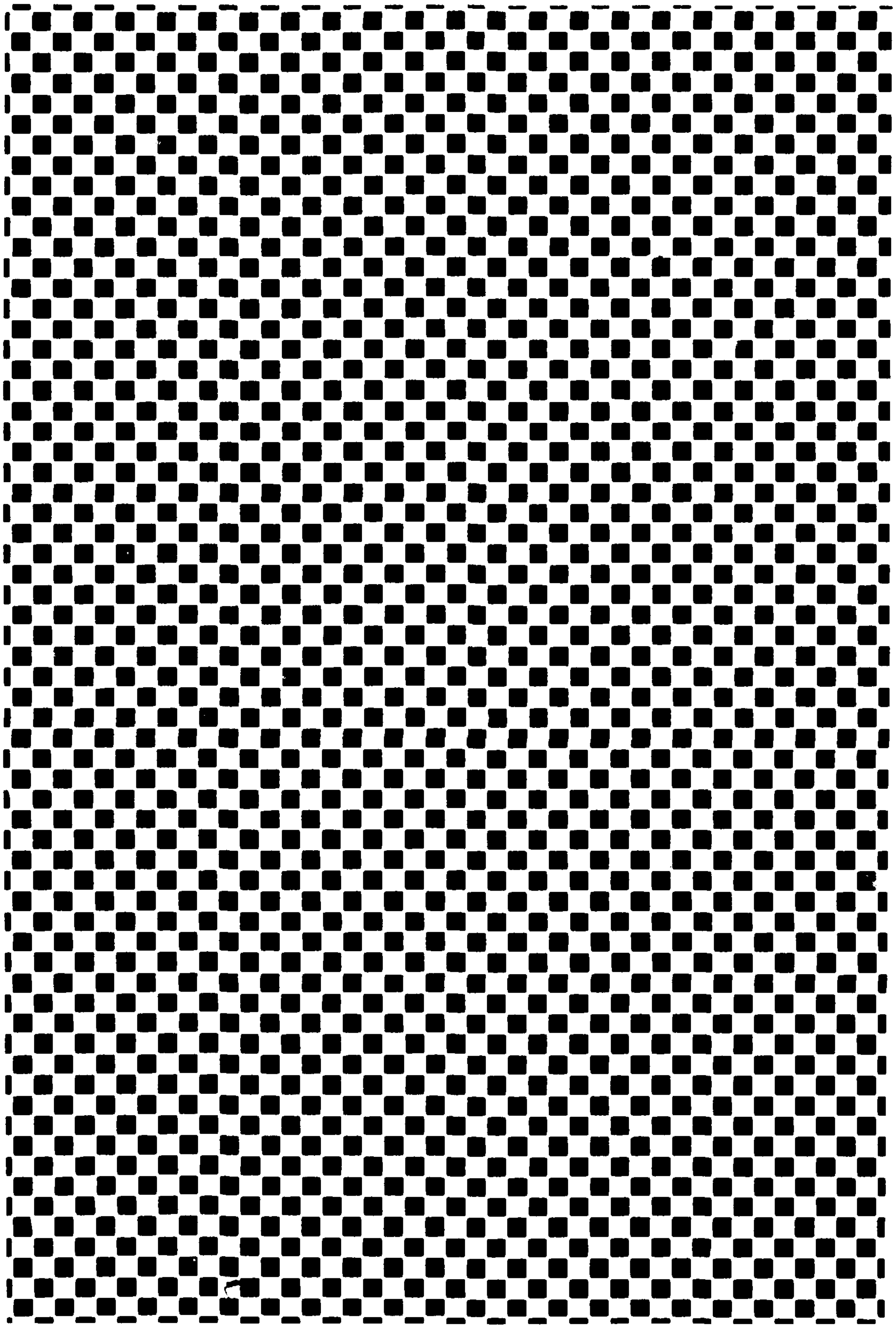


Figure 13. An example of a Checkerboard Target

Figure 14 gives the Z-score means for check 4. Windscreen B was judged to have the least distortion.

The discriminant analysis found that a linear combination of check 4 with check 1 produced excellent discrimination. The classification chart, shown in Table 10, indicates 86.3% correct classification. These two variables account for 99.7% of the variability in this study. A review of the plot of the first two canonical variates (Figure 15) shows that both variates are essential to discriminate between windscreens D and E and between A and B. The canonical correlations associated with these two variables were 0.993 and 0.769, respectively. Together, the two canonical variates accounted for 99.7% of the dispersion in the variables selected for inclusion in the discriminant analysis.

TABLE 10. CLASSIFICATION CHART FOR JUDGMENTS BASED ON CHECK 4 AND CHECK 1

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	2	1	0	0	0
B	1	2	0	0	0
C	0	0	3	0	0
D	0	0	0	3	0
E	0	0	0	0	3

A comparison of the plots of the first two canonical variates for the grids and checkerboards reveals some interesting differences. The first canonical variate for the checkerboard targets is more effective in separating the windscreens into three distinct groups than is the grid variate. Both target types discriminate; however, the second variate for the checkerboard seems especially useful in separating windscreen groups A and B. This experiment reaffirms the conclusion drawn from the grid study: two orthogonal dimensions are required to discriminate between windscreen groups A and B. It is evident that check 1 alone would not be a very efficient target since windscreens A and B have nearly equal ratings of distortion (see Figure 16). Rather it is the linear combination of the two variables that produces the best results.

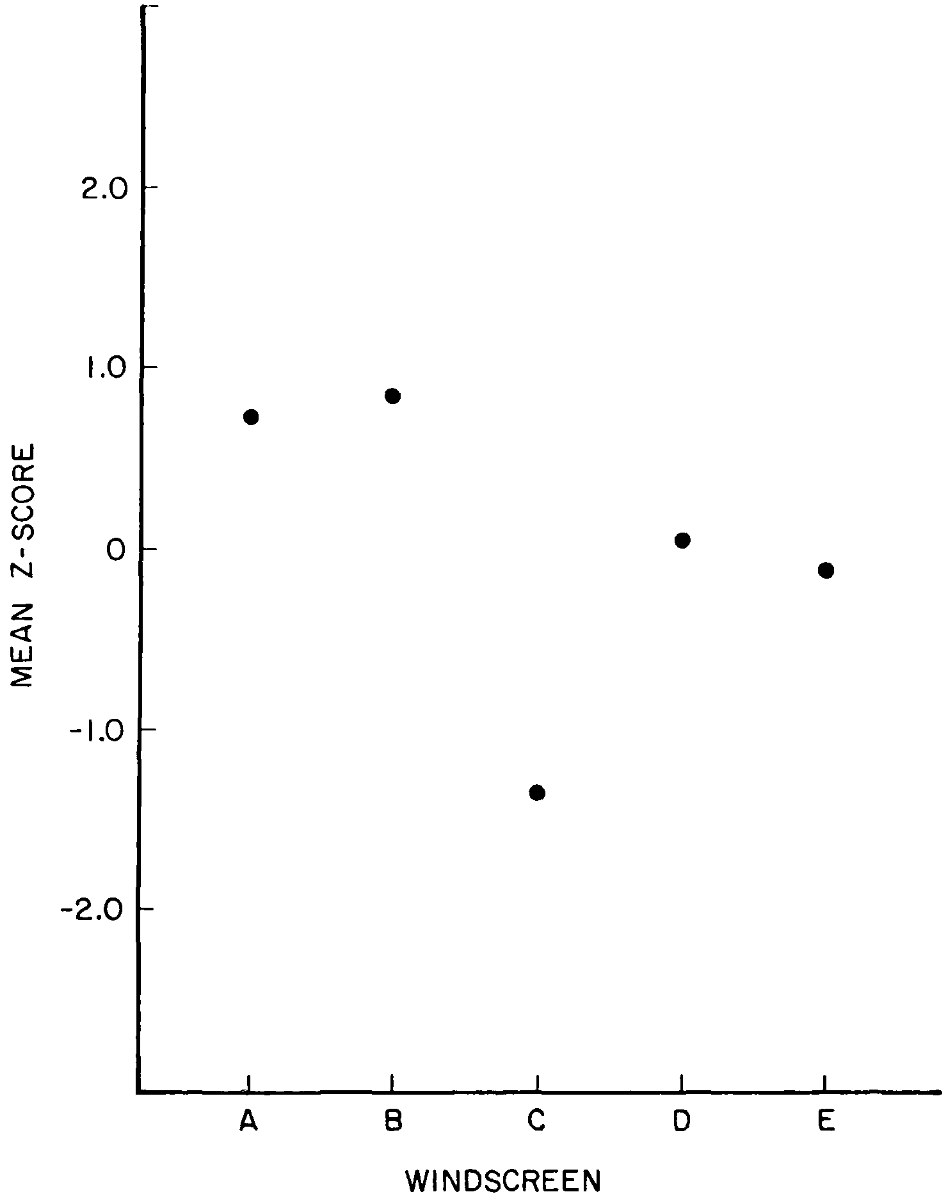


Figure 14. Z-Score Means (Judgments) with Check 4 as the Target

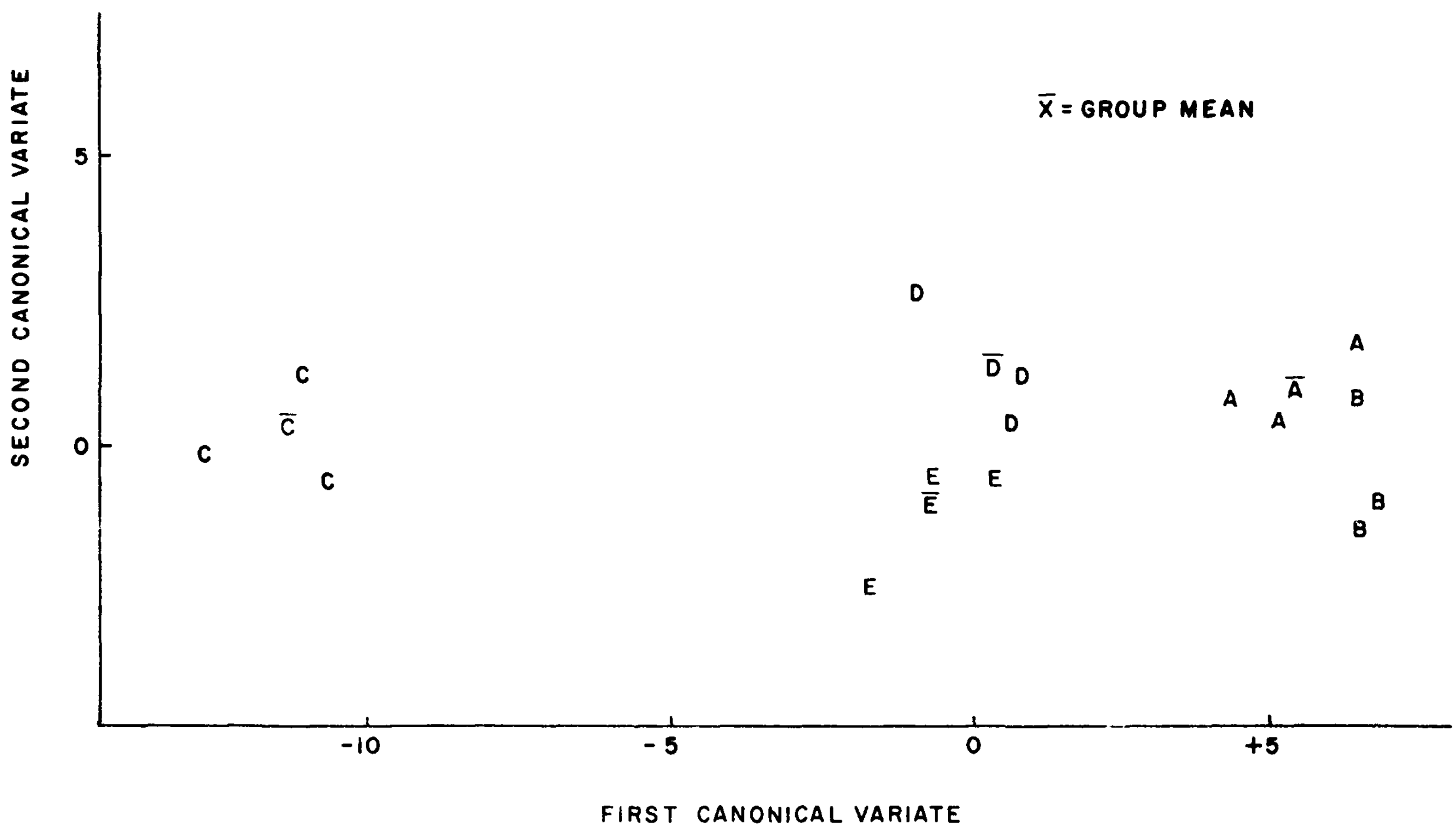


Figure 15. Plot of the First Two Canonical Variables for Experiment V

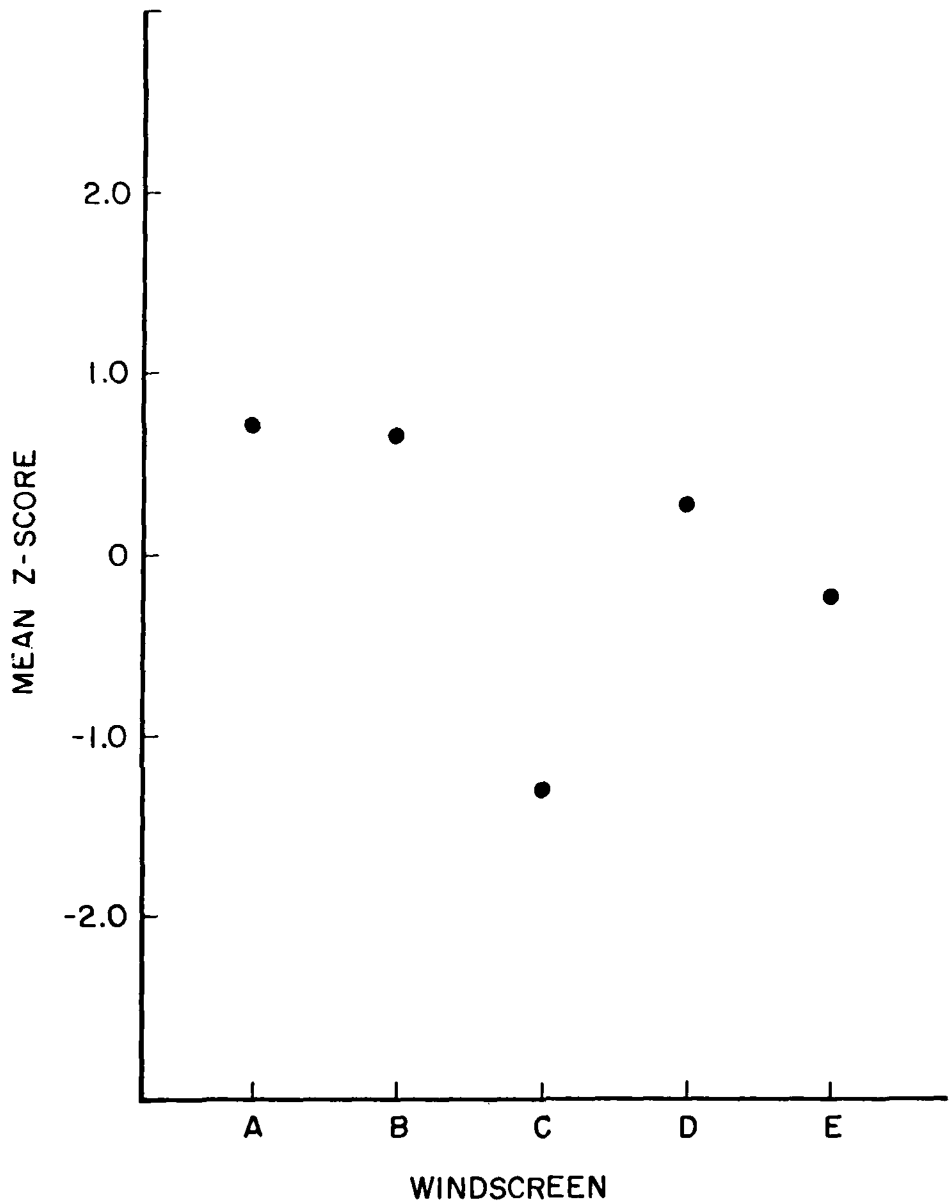


Figure 16. Z-Score Means (Judgments) with Check 1 as the Target

11. EXPERIMENT VI: COMPOSITE CHECKERBOARD AND GRID

This target was selected as a result of the analysis of the data from the grid and checkerboard targets. Since both target types seemed to enable subjects to judge distortion better than the others, it appeared reasonable to try to combine the grid with a checkerboard in the hope that each might capture some attribute of distortion that the other missed.

The data from the two types of targets were analytically combined in a discriminant analysis. That is, the data from all eight grids and four checks were analyzed together to see what linear combination would produce the best discrimination. In general, it appeared that the best targets were the largest checkerboard and the two smaller grids. Indeed it was possible to obtain 100% correct classification with the appropriate linear combination of a large checkerboard and small grids.

Since theoretical analysis worked well, two targets were fabricated using the largest checkerboard and superimposing either grid 1 or grid 2 over the white squares of the checkerboard. The dimensions of the targets are detailed in Table 11.

TABLE 11. DIMENSIONS OF COMPOSITE TARGETS

Composite	Checkerboard	Grid
1	0.37°	0.015 Black Lines 0.080 White Squares
2	0.37°	0.035 Black Lines 0.180 White Squares

The results for the best composite target are presented in Table 12. Composite target 2 produced the best discrimination with 60.0% correct windscreen classification.

TABLE 12. CLASSIFICATION CHART FOR JUDGMENTS BASED ON COMPOSITE 2

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	0	1	0	1	1
B	0	2	0	1	0
C	0	0	3	0	0
D	0	1	0	2	0
E	0	0	0	1	2

Figure 17 shows the Z-score means for the best composite target. From the figure, it is clear that windscreens A, B, and D are not very different from each other. Clearly, the use of the actual target in experiments did not produce the results that were predicted from the purely statistical analysis. It is possible that the two types of targets restrict responses when simultaneously present even though each captures an important aspect of distortion. Perhaps each pattern, when present by itself, allows the subject to respond to different characteristics of the windscreen in two separate measures. When observers are required to characterize a windscreen with one response, the information obtained is necessarily scaled unidimensionally. Thus, a single measure may not be as useful as those derived from multiple targets if the distortions are inherently multidimensional.

12. EXPERIMENT VII: GRATING

This study employed two square wave, horizontal bar targets. Each target was viewed both binocularly and monocularly. Table 13 outlines the grating sizes and viewing conditions.

TABLE 13. GRATING TARGET SIZES AND VIEWING CONDITIONS

Grating	Bar Size in Degrees of Visual Angle	Viewing Condition
1	0.15	Binocular
2	0.30	Binocular
3	0.15	Right Eye Only
4	0.30	Right Eye Only
5	0.15	Left Eye Only
6	0.30	Left Eye Only

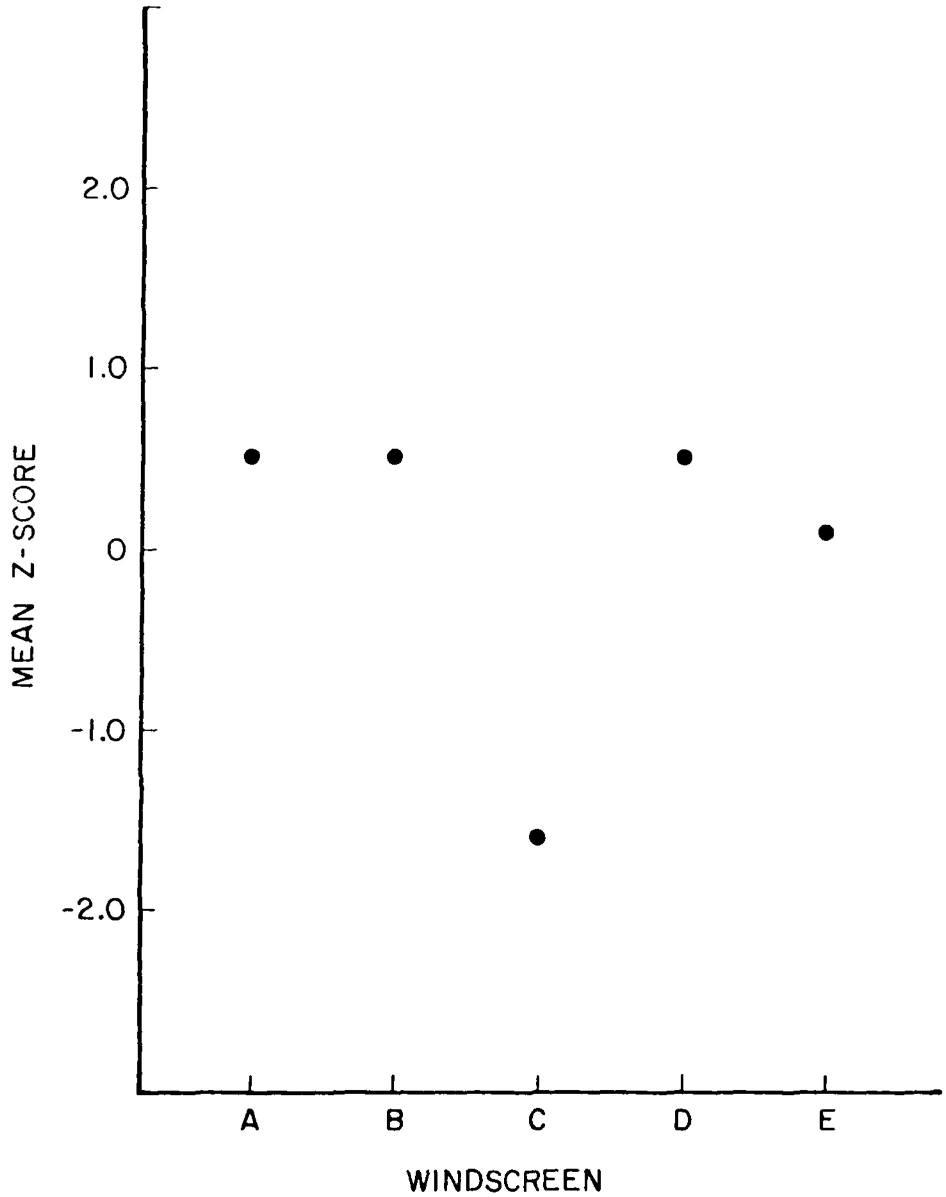


Figure 17. Z-Score Means (Judgments) Using Composite Target 2

The grating targets were drawn and photographed for projection as 2 x 2-inch slides. A suitable lens was chosen for the projector such that the 8 x 8-foot screen was filled by the projected target. Center field luminance was approximately 5.4 footlamberts with 48 percent falloff near the edges. (The 2 x 2 projection clearly did not have the optical quality of the lantern slide projector.) Target density was 0.28 and the black bars had a contrast of 0.94.

Three subjects viewed the targets through the windscreens and judged distortion as before. The discriminant analysis showed that the "best" target was the 0.15° grating combined with the monocular (right eye) viewing condition. The classification chart, shown in Table 14, reveals that 60.0% of the windscreens could be correctly classified with the use of this target.

TABLE 14. CLASSIFICATION CHART FOR JUDGMENTS BASED ON GRATING 3

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	2	1	0	0	0
B	0	2	0	1	0
C	0	0	3	0	0
D	0	2	0	1	0
E	1	0	0	1	1

In general, windscreens A and B tended to be confused with E and D, respectively. The Z-scope means, shown in Figure 18, indicate why this occurred. The means for the two groups were very similar.

The discriminant analysis extracted a better solution when a linear combination of two variables was formed. As is shown in Table 15, when the data from grating 1 were combined with that of grating 3, 80.0% correct classification was possible. The A-E and B-D confusions were largely eliminated.

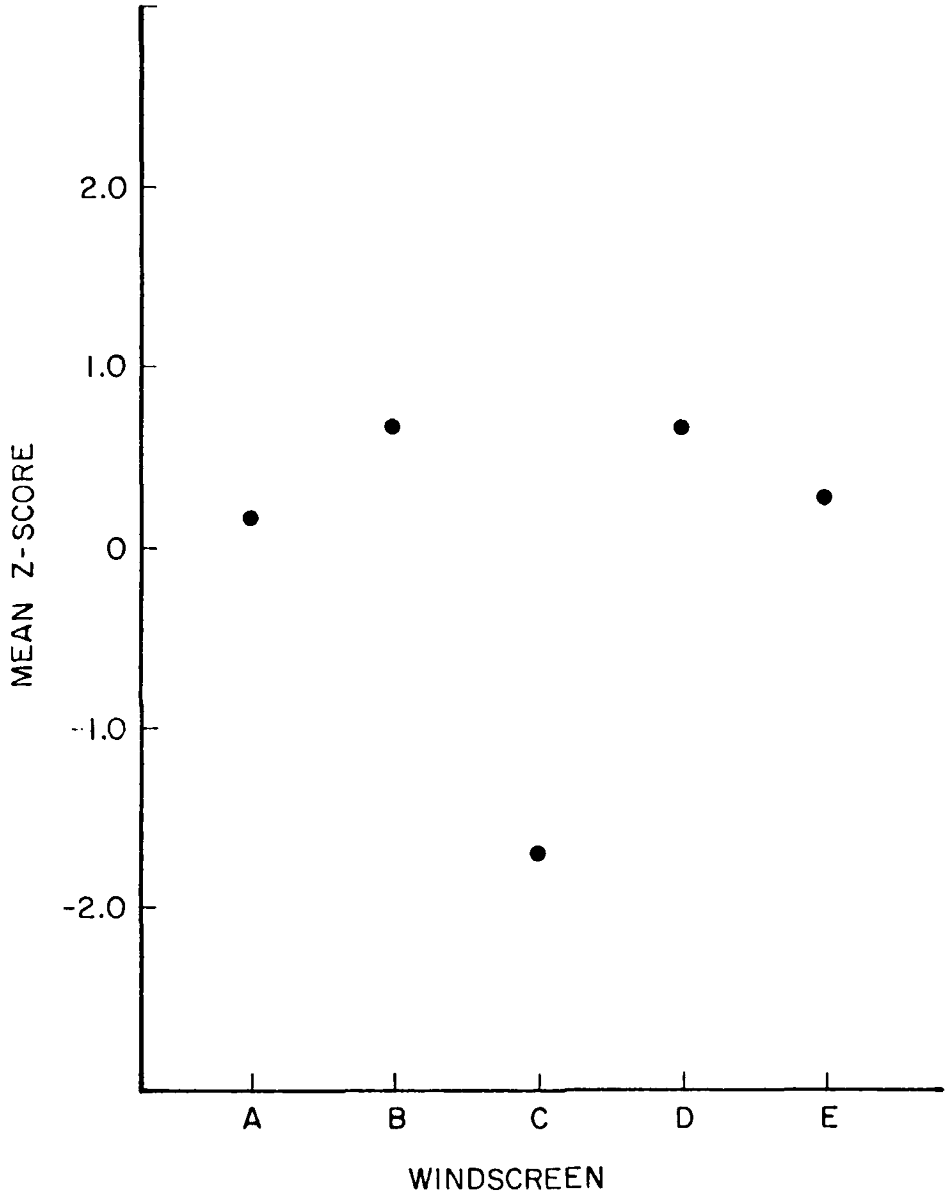


Figure 18. Z-Score Means (Judgments) with Grating 3 as the Target

TABLE 15. CLASSIFICATION CHART FOR JUDGMENTS BASED ON GRATING 3 AND GRATING 1

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	2	0	0	1	0
B	0	3	0	0	0
C	0	0	3	0	0
D	0	1	0	2	0
E	0	0	0	1	2

Figure 19 shows the means obtained when grating 1 was used as the target.

The results of the grating experiments show that not only does the target significantly influence judgments of windscreen distribution, but also that the nature of the viewing condition influences the judgments of windscreen quality. The "best" single variable involved monocular viewing. When a linear combination of two variables was formed, both monocular and binocular viewing contributed to the discrimination. Because canonical variates are mutually orthogonal, it appears that binocular viewing must enable subjects to utilize some additional aspect of distortion that is not available when monocular vision is employed. This conclusion is based on the analytic solution. The perceptual consequences of the viewing conditions are not immediately obvious.

Binocular viewing requires subjects to visually sample the target arrays from two different spatial locations. Often, distortion is different for each spatial location, so that the elements of the grating target stimulate noncorresponding retinal areas. This noncorrespondence must be resolved by the viewer's perceptual system. Normally, small, systematic noncorrespondence is perceived as depth, but the irregular pattern of distortion does not induce such effects. There do not appear to be any easily identified perceptual phenomena associated with the binocular viewing condition. One's subjective impression is that distortion is enhanced, although it is difficult to describe the perceptual differences between the two viewing modalities.

A plot of the two canonical variates (Figure 20) shows that most of the groups of windscreens are well-separated. The canonical correlations for the

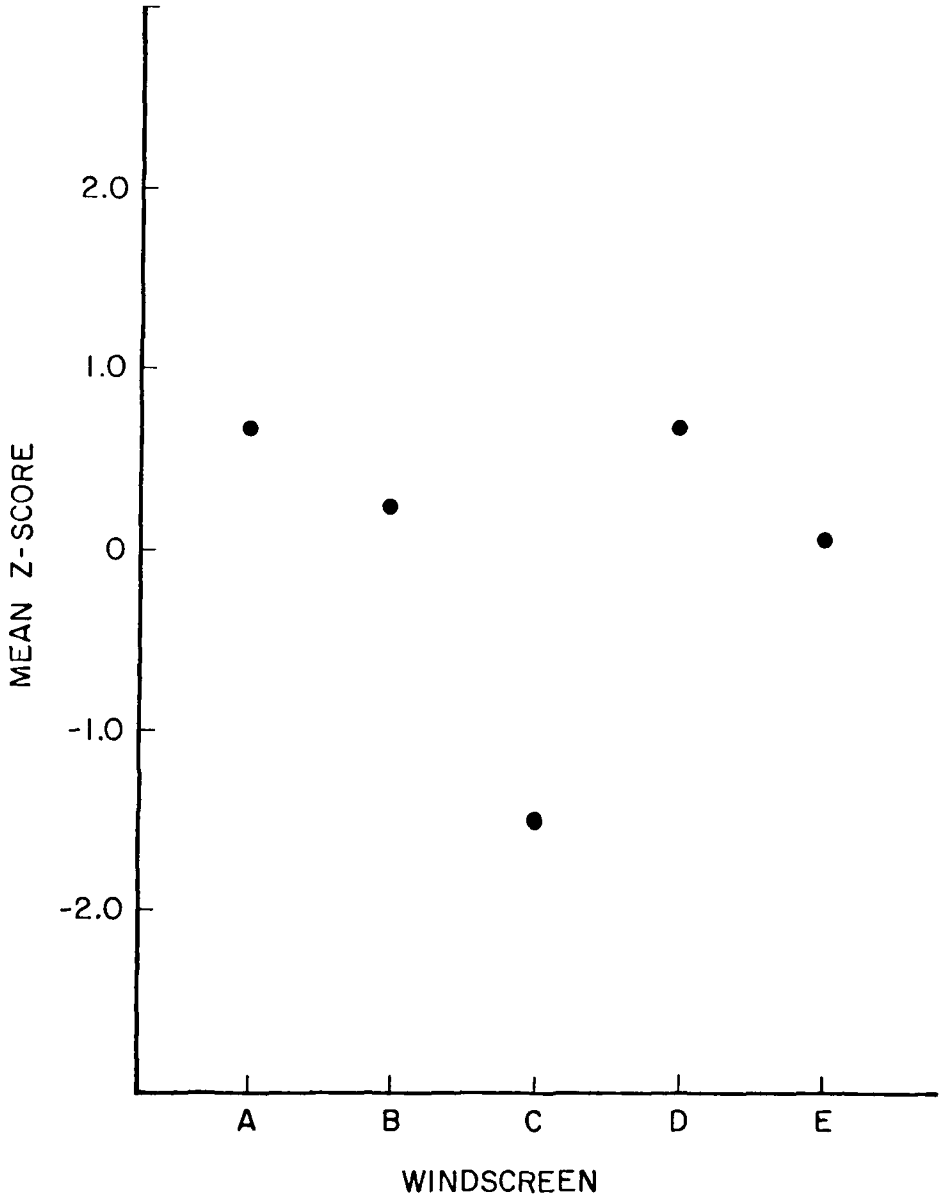


Figure 19. Z-Score Means (Judgments) with Grating 1 as the Target

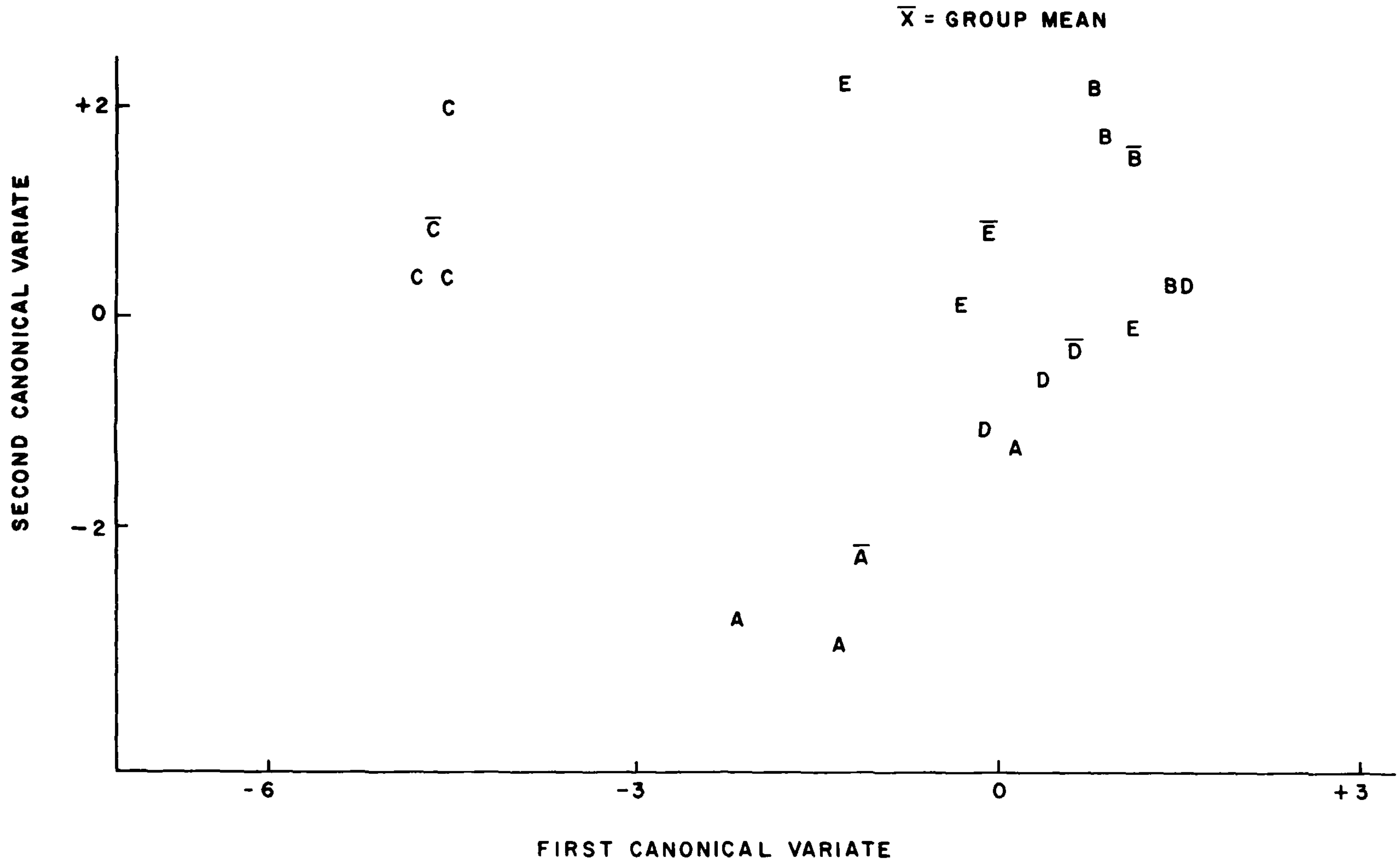


Figure 20. Plot of the First Two Canonical Variates for Experiment VII

first and second variates were 0.944 and 0.862 respectively. Both of these correlations are quite high, indicating that subjects' judgments and windscreens are highly related for both variates. Clearly, the second variate contributes information necessary for discrimination.

An examination of the F-statistic for the grating conditions indicates that the large grating is not a particularly good target. After grating condition 3 was included in the discriminant function, all of the remaining small grid variables had higher F values than the large gratings.

13. EXPERIMENT VIII: GRATING, SINUSOIDAL MOVEMENT

This experiment and the next two describe investigations wherein dynamic targets were used. Many distortions appeared more pronounced if the observer moved his head slightly. Parallax changes enhanced distortions and, according to subjects' reports, made judgments easier. In the previous experiments head motion was prevented by requiring subjects to use a chin rest. In the dynamic target studies, target motion was intentionally introduced so that the motion could be controlled and quantified, while the subject's head remained fixed.

In this study, six subjects viewed the same two gratings that were used in Experiment VII, except that the projector was mounted on a stand that oscillated in response to the driving waveform of a signal generator. The target gratings moved vertically with a sinusoidal velocity function that averaged either 2°/sec or 3°/sec. The grating conditions are outlined in Table 16. Subjects were instructed to wait until the target had cycled up and down a few times before making their responses (see Appendix A).

TABLE 16. SINUSOIDAL MOTION GRATING CONDITIONS

Grating	Bar Size in Degrees Visual Angle	Viewing Condition	Average Speed in Degrees/Sec
1	0.30	Binocular	2
2	0.30	Binocular	3
3	0.15	Binocular	2
4	0.15	Binocular	3
5	0.30	Right Eye	2
6	0.30	Right Eye	3
7	0.15	Right Eye	2
8	0.15	Right Eye	3
9	0.30	Left Eye	2
10	0.30	Left Eye	3
11	0.15	Left Eye	2
12	0.15	Left Eye	3

The results of this study are shown in Table 17. The "best" variable (target) was grating 8 which led to 66.7% correct classification.

TABLE 17. CLASSIFICATION CHART FOR JUDGMENTS BASED ON GRATING 8

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	4	0	0	0	2
B	0	4	0	2	0
C	1	0	5	0	0
D	0	1	0	5	0
E	3	1	0	0	2

Mean Z-scores obtained when grating 8 was the target are presented in Figure 21. As in the previous grating study, a better classification was produced by a linear combination of variables. Table 18 shows the classification obtained when the data from grating 8 and 1 were analytically combined. This combination yielded 83.3% correct classification.

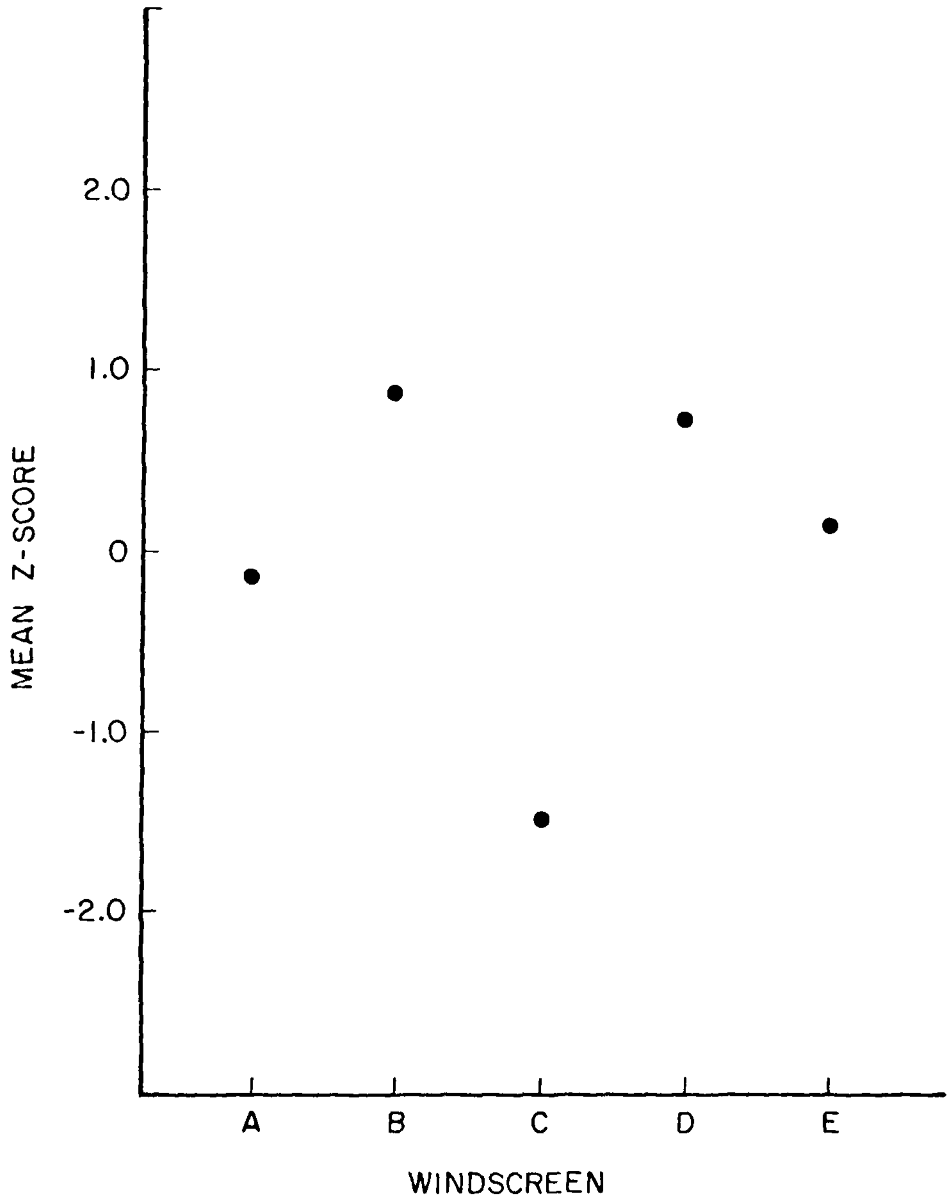


Figure 21. Z-Score Means (Judgments) with Sinusoidal Motion, Grating 8 as the Target

TABLE 18. CLASSIFICATION CHART FOR JUDGMENTS BASED ON GRATING 8 AND GRATING 1

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	5	0	0	0	1
B	0	4	0	2	0
C	0	0	6	0	0
D	0	1	0	5	0
E	0	1	0	0	5

Figure 22 shows Z-score means for grating 1.

The combination of the two variables accounted for 92.7% of the total variability in the data. Actually the analysis revealed that 100% discrimination could be achieved, although this required using nine of the twelve variables! From a practical standpoint, two variables appeared to capture the essence of the results. Figure 23 shows the plot of the first two canonical variates. The canonical correlations associated with the first and second variates were 0.987 and 0.937 respectively. Inspection of the plot shows that, while windscreen groups C and A were well separated, groups B, D, and E were somewhat mixed. It is interesting to note the parallels between the study with static gratings and the sinusoidal movement experiment. The best variable in both experiments was the monocular viewing condition for the grating with the thin lines. The binocular condition was chosen as the second variable to be included in the discriminate analysis. Although the 0.30° grating appeared to be a slightly better target in the binocular condition of the dynamic study, it should be noted that the third step in the discriminate analysis included grating condition 4. Thus, the two grating studies are actually in close agreement.

14. EXPERIMENT IX: GRATING, LINEAR MOVEMENT

This work extends the previous study to include linear motion of target gratings. Several subjects in the previous study complained about the rapid instantaneous velocity in the center of the target excursion that was due to the sinusoidal velocity profile. They pointed out that the rapid motion made judgments very difficult. The present study was therefore designed to utilize

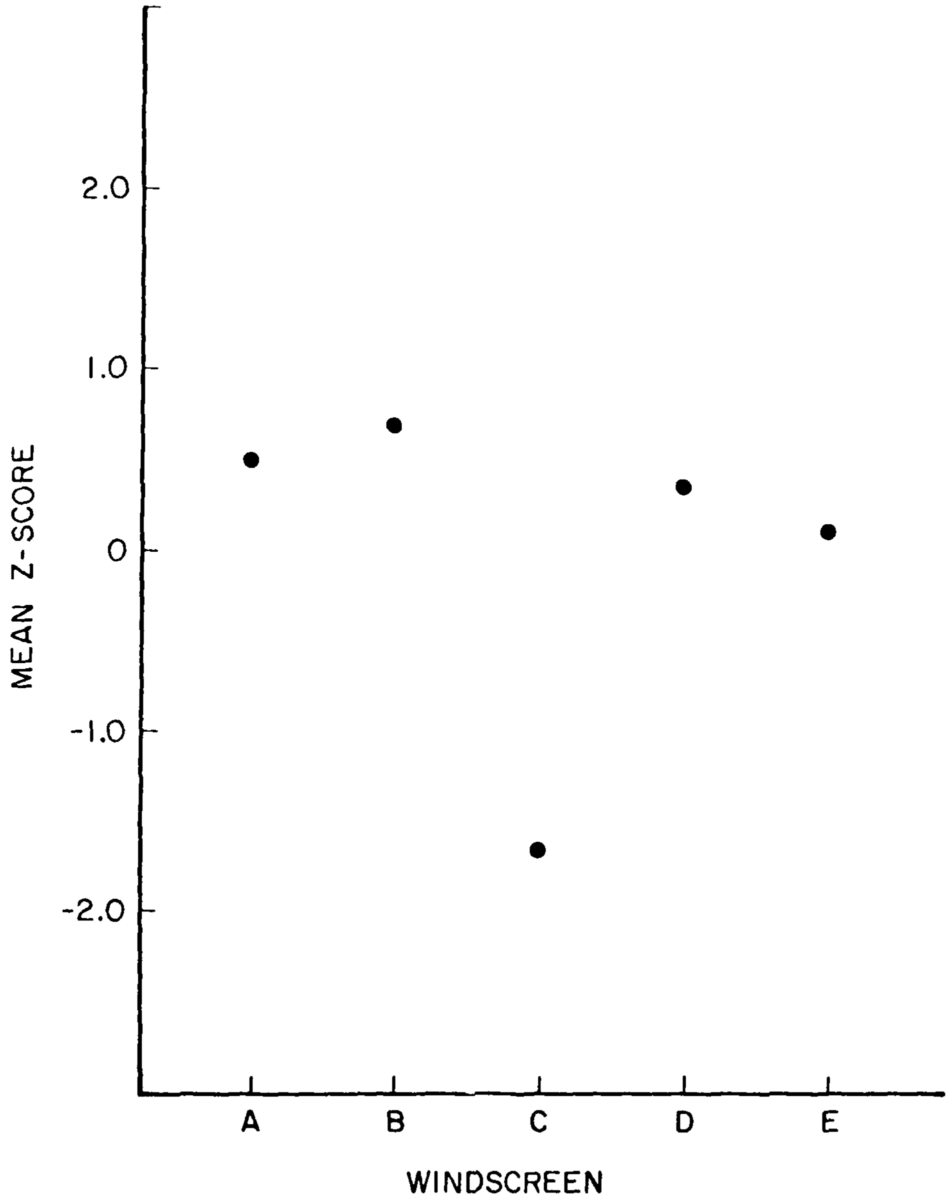


Figure 22. Z-Score Means (Judgments) with Sinusoidal Motion, Grating 1 as the Target

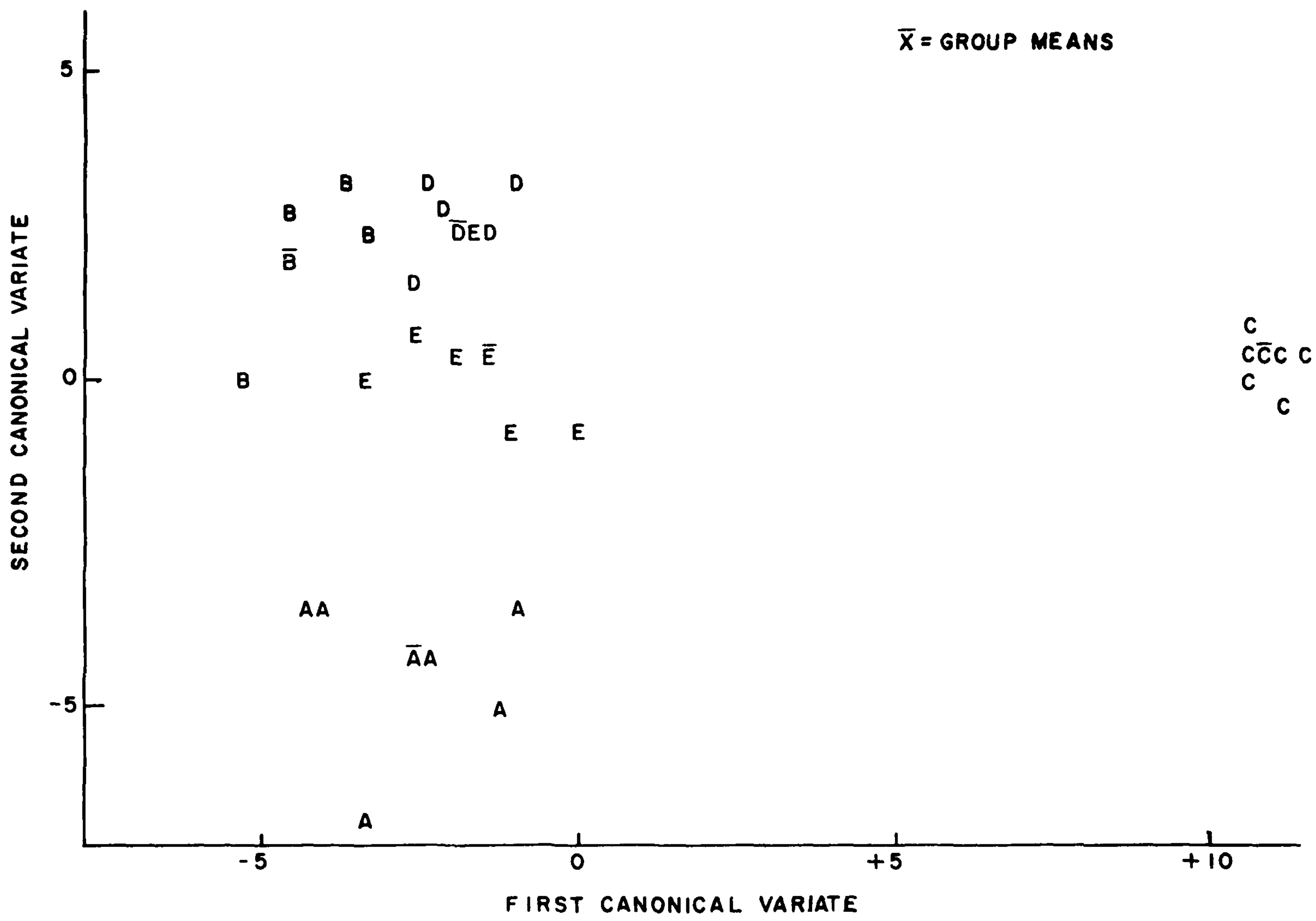


Figure 23. Plot of the First Two Canonical Variates for Experiment VIII

linear motion to determine if the velocity profile was an important aspect of making judgments about windscreen distortion. Only the 2°/sec velocity was used because subjects had complained that the fast average velocity was a problem.

The conditions of Experiment IX are outlined in Table 19. Binocular viewing was used and four subjects participated.

TABLE 19. LINEAR MOTION GRATING CONDITIONS

Grating	Bar Size In Degrees Visual Angle
1	0.15
2	0.30

The results of the linear motion experiment are summarized in Table 20. The analysis indicated that only grating 1 was statistically significant for classifying windscreens. Correct classification of 60% was obtained with this variable.

TABLE 20. CLASSIFICATION CHART FOR JUDGMENTS BASED ON GRATING 1

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	2	0	0	1	1
B	0	4	0	0	0
C	0	1	3	0	0
D	0	1	0	1	2
E	2	0	0	0	2

Figure 24 shows the Z-score means for grating 1. Clearly, the range of judgments was compressed for this variable compared with the data of grating 8 (Figure 21) and grating 1 (Figure 22) in the previous study. This effect is reflected by the difference in the F-statistics generated by the multivariate analysis of variance that precedes the discriminant analysis. The sinusoidal movement study produced F values of 58.27 and 53.53 for gratings 8 and 1, respectively, whereas linear motion only resulted in an F value of

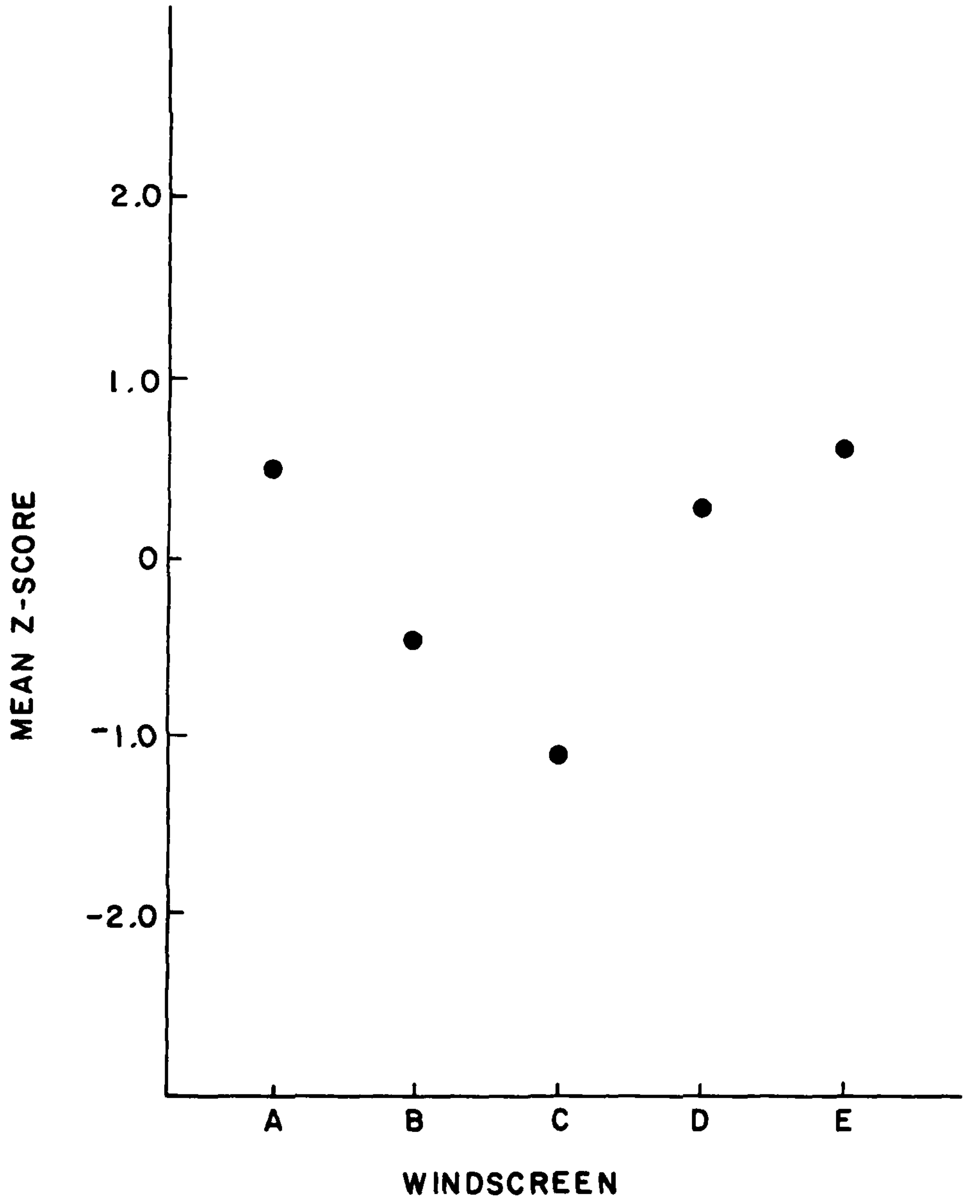


Figure 24. Z-Score Means (Judgments) with Linear Motion, Grating 1 as the Target

13.31 for grating 1*. The analysis indicates that even though the percent classification for the present experiment (60%) is not very different from that produced by grating 8 in the previous work (67.7%), the effect of sinusoidal motion is really much stronger than linear motion. The reason for this is not clear and contrary to expectation.

When the binocular conditions of the two experiments are compared, it is obvious that the sinusoidal motion condition is superior to the linear motion one. Thus, even though subjects complained about the rapid instantaneous velocity of target movement in Experiment VIII, that study demonstrated that valid judgments could be made with this arrangement.

When the data of the three grating studies are considered together, some interesting conclusions can be drawn. First, for overall windscreen classification, according to the perceived severity of distortion, the static grating appears to be nearly as effective as one that undergoes sinusoidal movement. Linear motion, coupled with a grating target, on the other hand, is not a good viewing condition for stimulating accurate judgments of distortion found in aircraft windscreens.

Second, while the sine-motion and static grating produce about the same percent correct classification, examination of the plots of the canonical variates shows that the pattern is different. Windscreens B, D, and E each tend to be grouped together in Figure 23. For the static grating, however, group E has such high variability that it is grouped with both D and B.

Third, the grating target is most effective when judgments are made monocularly and then with binocular viewing. Maximal discrimination among windscreen groups is possible only when a linear combination is formed from both monocular and binocular variables.

---

*The degrees of freedom are slightly different for the two experiments, but even allowing for this, there is a dramatic difference in the significance levels of the two studies.

15. EXPERIMENT X: LINEAR MOTION BAR

This study was similar to Experiment IX* with one exception. Instead of using a grating, only a single horizontal line or bar (having contrast of 0.88) was presented. It was moved in a linear movement pattern up and down on the target screen at a speed of 2°/sec. Four subjects were used and were instructed as before. The target configurations are outlined in Table 21.

TABLE 21. BAR TARGET CONDITIONS

Bar	Bar Size in Degrees Visual Angle
1	0.15
2	0.30

The results of the single bar target show that of the two bar sizes, bar 1 was better in determining windscreen classification. These data are presented in Table 22. The bar 1 variable produced 60.0% correct classification.

TABLE 22. CLASSIFICATION CHART FOR JUDGMENTS BASED ON BAR 1

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	3	0	0	1	0
B	0	3	0	1	0
C	0	0	4	0	0
D	1	0	0	2	1
E	2	2	0	0	0

The mean Z-scores obtained with bar 1 as the target are shown in Figure 25.

The analysis revealed that better classification could be obtained with a linear combination of the variables. When bar 1 and bar 2 data were combined, 85.0% correct classification was possible, as shown in Table 23.

*This experiment was planned and under execution before the data of Experiments VIII and IX were analyzed. Hence we were unaware of the apparent superiority of sinusoidal movement.

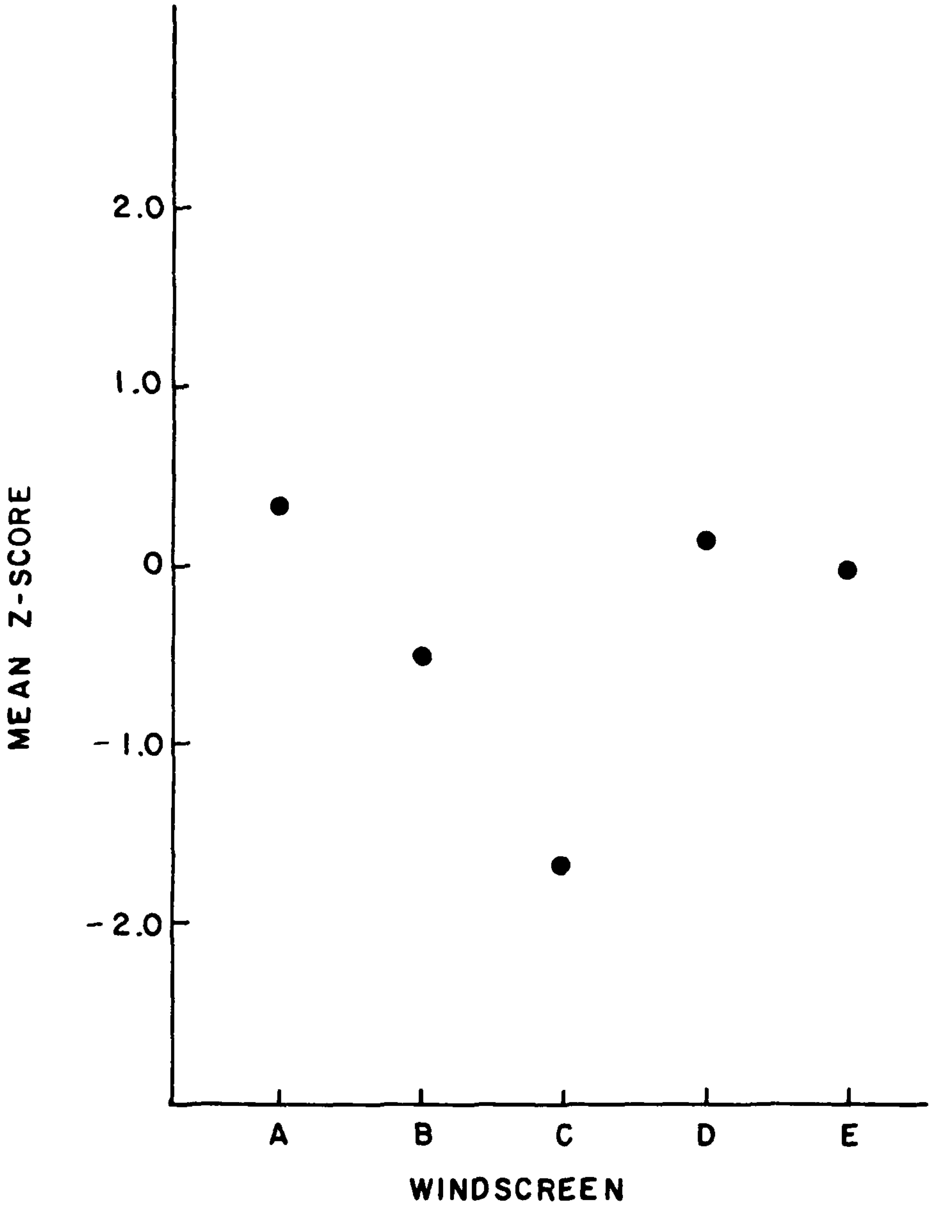


Figure 25. Z-Score Means (Judgments) with Linear Motion, Bar 1 as the Target

TABLE 23. CLASSIFICATION CHART FOR JUDGMENTS BASED ON BAR 1 AND BAR 2

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	3	0	0	1	0
B	0	3	0	1	0
C	0	0	4	0	0
D	1	0	0	3	0
E	0	0	0	0	4

Figure 26 shows the Z-score means when bar 2 is used on the target.

Adding the second variable contributed greatly to the discrimination. Figure 27 shows the plot of the two canonical variates for the bar target. The canonical correlations were 0.919 and 0.869 for the first and second variates, respectively. These high correlations show that both variates are important in the discrimination. A review of the plot shows that both variates are required for good discrimination.

16. CONCLUSIONS, PHASE I

Confusability among distortion judgments for windscreen groups appears to be a function of target type. The data from Experiment X, Linear Motion of a Bar Target, show that windscreen groups A and D are the most difficult to separate. This is a different type of group confusion than is evident in the moving grating experiments. In those studies, groups B and D (Experiment VIII) and A and E (Experiment IX) tended to be the most difficult to discriminate. Groups A and B are the least discriminable in Experiment I, the grid study.

Although a variety of targets provide good discrimination among windscreens, the differential confusability among groups leads to the conclusion that each target may highlight different "aspects" of distortion. These aspects or features of distortion cannot be identified at present because very little is known about the perception of distortion. Certain target types do appear to facilitate distortion judgments and suggest the following conclusions:

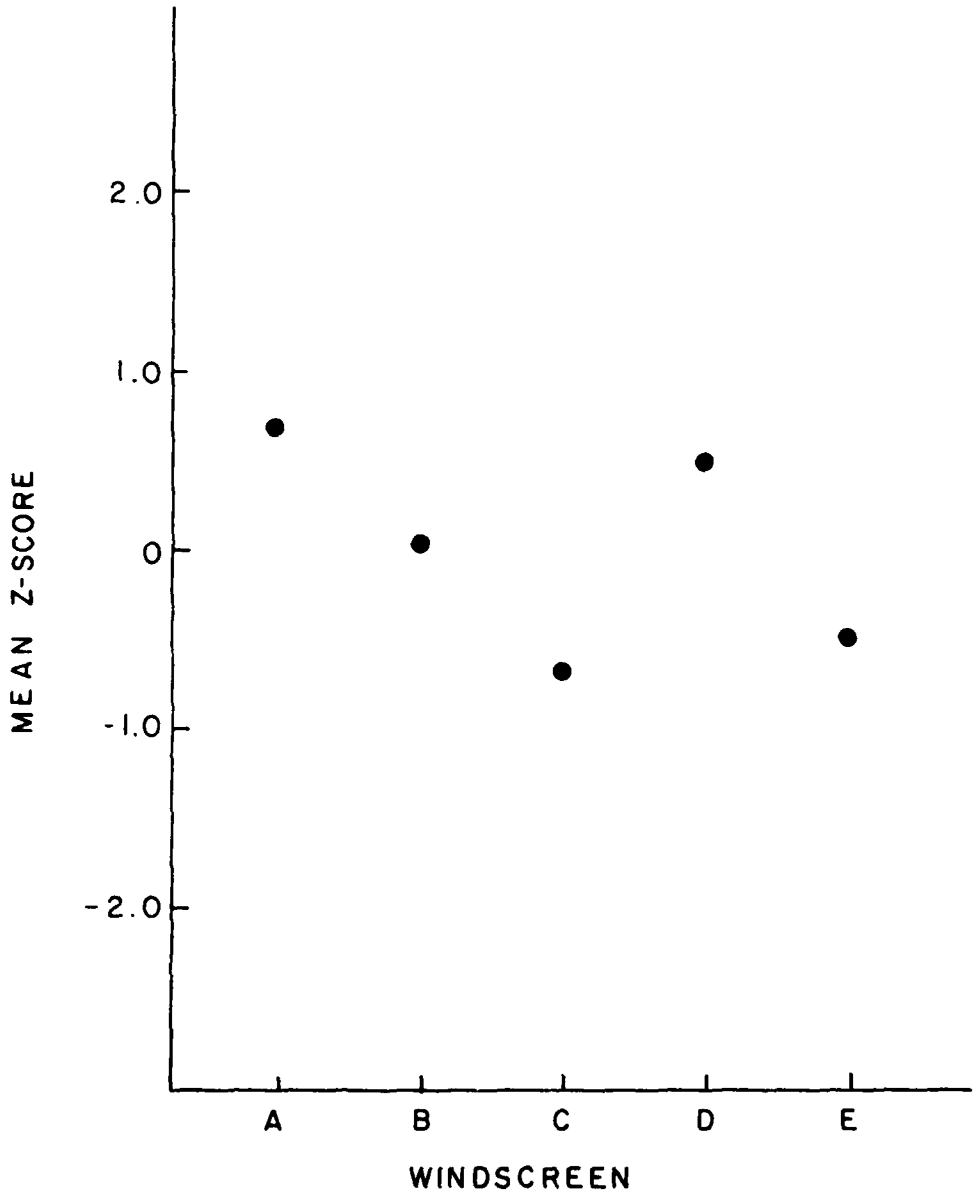


Figure 26. Z-Score Means (Judgments) with Linear Motion, Bar 2 as the Target

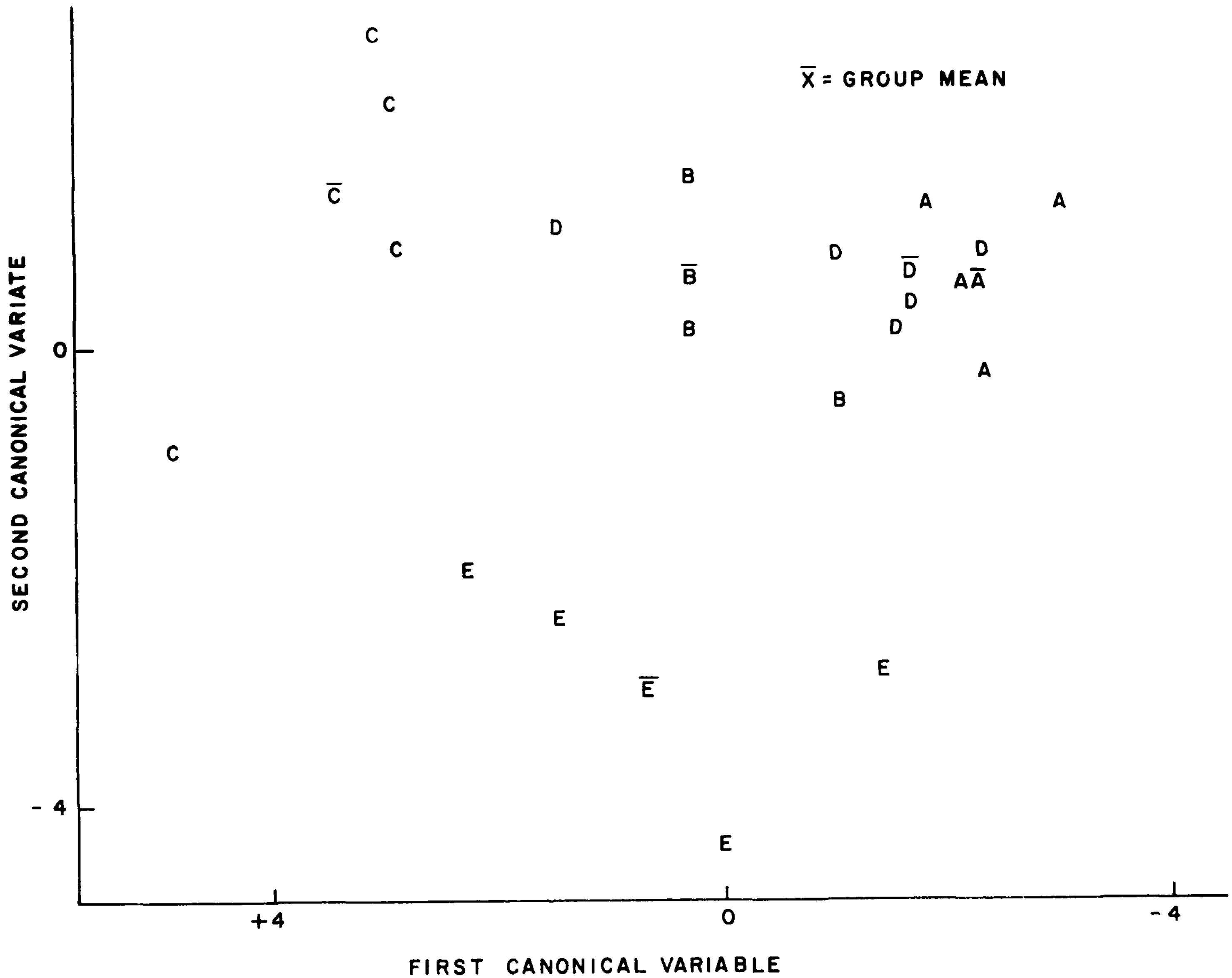


Figure 27. Plot of the Two Canonical Variates for Experiment X

(1) Targets having rectilinear or periodic line elements will be superior to aperiodic targets that do not have linear arrays.

(2) Dynamic targets are appropriate if suitable motion parameters are chosen.

(3) The most promising studies should be replicated with a larger sample of subjects.

SECTION III

PHASE II EXPERIMENTS

1. APPROACH

The Phase II work required that the "best" targets studied in Phase I be selected and the studies conducted again using a larger number of subjects and the most promising conditions. Our decision in selecting Phase II targets was based principally upon the windscreen classification data. These Phase I data have been tabulated and are presented together in Table 24 (page 67) so the comparisons can be made easily. Note that both the grid study and the grating (with sinusoidal movement) work used six subjects. Both of these studies showed that the targets and conditions were appropriate for windscreen classification. However, the best overall classification resulted when checkerboards and a bar target (with linear motion) were used. Since these latter two studies employed only three and four subjects, respectively, it seemed reasonable to select them for replication with a larger number of subjects.

2. EXPERIMENT XI: CHECKERBOARD

In this study seven subjects were run, using the two best variables that emerged from the Phase I checkerboard study. These targets and their dimensions are outlined in Table 25.

TABLE 25. PHASE II CHECKERBOARD TARGETS

Check	Size in Degrees Visual Angle
1	0.08
2	0.37

Check 2 was the first variable chosen for inclusion in the discriminant analysis, It permitted 51% of the windscreens to be correctly classified. Table 26 shows the results of the classification. Clearly, discrimination was poor for this target and worse than in Phase I.

TABLE 24. WINDSCREEN CLASSIFICATIONS FOR PHASE I EXPERIMENTS

	Experiment	Number of Subjects	Classification by One Variable, Percent Correct	Classification* by Two Variables, Percent Correct
I.	Grids	6	73	80
II.	Periodic Dot	3	53	
III.	Texture	3	53	
IV.	Meridional	3	47	
V.	Checkerboard	3	73	86
VI.	Composite Check and Grid	3	60	
VII.	Grating, Static	3	60	80
VIII.	Grating, Sine Movement	6	67	83
IX.	Grating, Linear Movement	3	60	
X.	Bar, Linear Movement	4	60	85

*Some experiments have only one variable (target type) and therefore have only one basis for classification. Experiments that utilized two or more targets may have two of these variables involved in the classifications reported for the second step of the discriminate analysis. Inclusion of two variables depends upon the F-statistic prevailing at the second step of the analysis. In cases where the analysis proceeded beyond two steps, only the classification in effect at the second step has been reported. This limit was imposed because most of the information useful for discrimination is present in two canonical variates. (The inclusion of three or more target variables in these first two variates usually added very little to group discrimination.)

TABLE 26. CLASSIFICATION CHART FOR JUDGMENTS BASED ON CHECK 2

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	4	1	0	1	1
B	0	5	1	1	0
C	1	1	5	0	0
D	2	2	1	2	0
E	3	1	0	1	2

The second variable, check 1, was also significant and the classification based on a linear combination of the two variables is shown in Table 27. Again, only 51% of the windscreens were correctly classified. Figures 28 and 29 show the Z-score means for checks 2 and 1 respectively. When the results of this study are compared with those of Experiment V, it is obvious that this second checkerboard study failed to replicate the previous findings.

TABLE 27. CLASSIFICATION CHART FOR JUDGMENTS BASED ON CHECK 1 AND CHECK 2, SEVEN SUBJECTS

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	1	1	0	2	3
B	0	6	0	1	0
C	0	0	5	1	1
D	2	1	1	2	1
E	2	0	0	1	4

Since many new, inexperienced subjects were recruited for the Phase II work, it was decided to examine subjects' judgments in detail. The frequency of judged score intervals was plotted for each target, windscreen, and subject for both checkerboard studies. In general, most subjects in both studies seemed to make similar assessments of windscreen quality. Two subjects in Phase II were dropped from the analysis because they failed to differentiate among the windscreens and always judged the distortion as greater than the rest of the observers, regardless of the windscreen presented. It was therefore decided to combine the relevant data from the two checkerboard

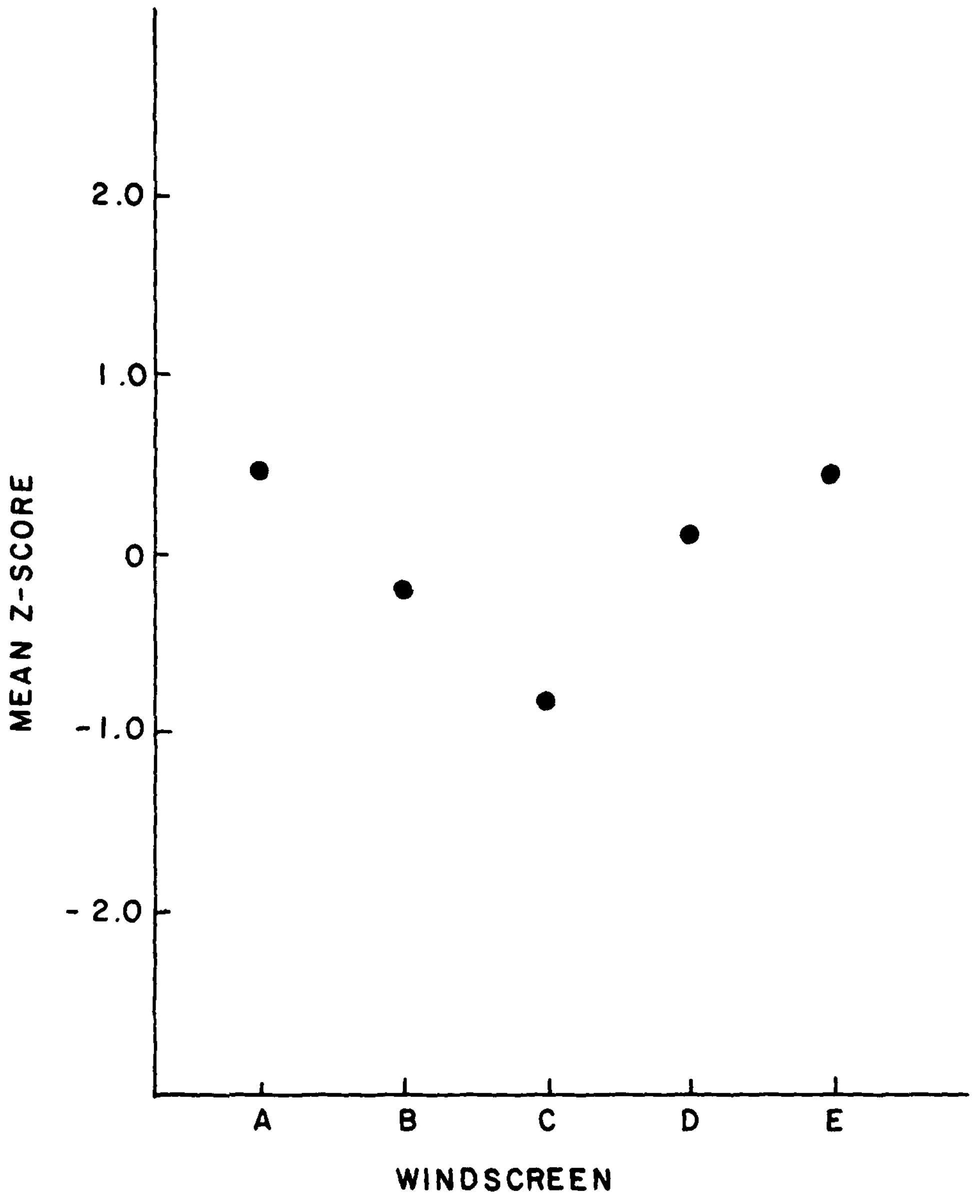


Figure 28. Z-Score Means (Judgments) with Checkerboard 2 as the Target

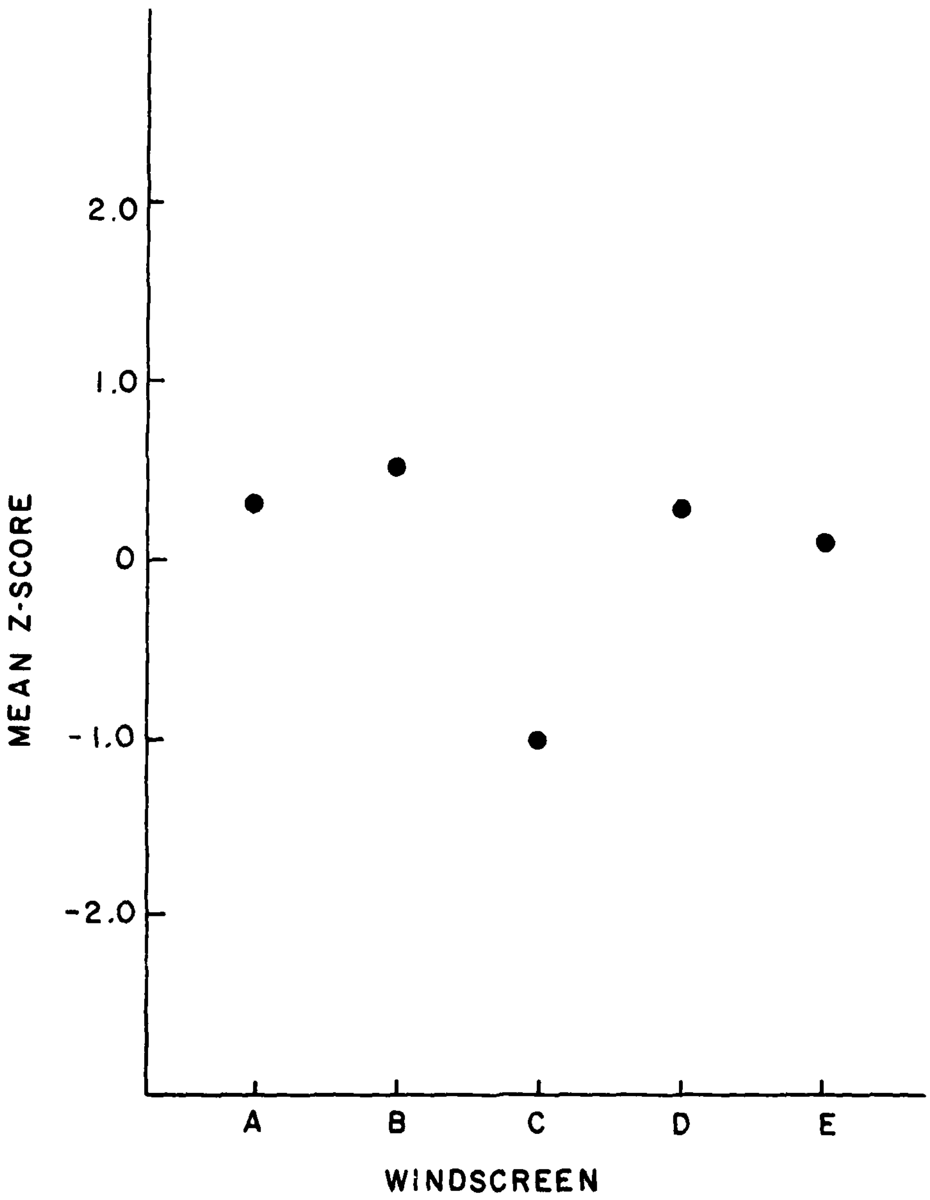


Figure 29. Z-Score Means (Judgments) with Checkerboard 1 as the Target

studies and analyze the data for the remaining eight subjects for check 1 (0.08°) and check 2 (0.37°).

The analysis based on the eight subjects showed that both checkerboard targets contributed significantly to windscreen discrimination. The classification chart for the linear combination of the two targets is shown in Table 28.

TABLE 28. CLASSIFICATION CHART FOR JUDGMENTS BASED ON CHECK 1 AND CHECK 2, EIGHT SUBJECTS

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	3	2	0	1	2
B	3	4	0	0	1
C	0	0	8	0	0
D	2	2	0	2	2
E	3	1	0	0	4

Removal of the eccentric data did not appreciably improve classification. Only 57.5% correct classification was achieved. The plot (Figure 30) of the two canonical variates shows that, except for windscreen C, no clear grouping is evident.

Further analysis of the data of Phase II revealed that the seven subjects participating in this phase had more intersubject variability than those of Phase I. One example of this is illustrated in Figure 31, which shows Z-score means and standard deviations for Phase I and Phase II subjects for windscreen A. Judgments for other windscreens showed similar effects. Subjects 1-3 (Phase I) were much more consistent than subjects 4-10 (Phase II).

Overall, the less experienced subjects were not as consistent as those observers who participated in the first checkerboard study. For illustration, the standard deviations for check 1 and check 2 are shown in Table 29 for the Phase I and Phase II work, and for the last analysis based on three Phase I subjects and five Phase II subjects.

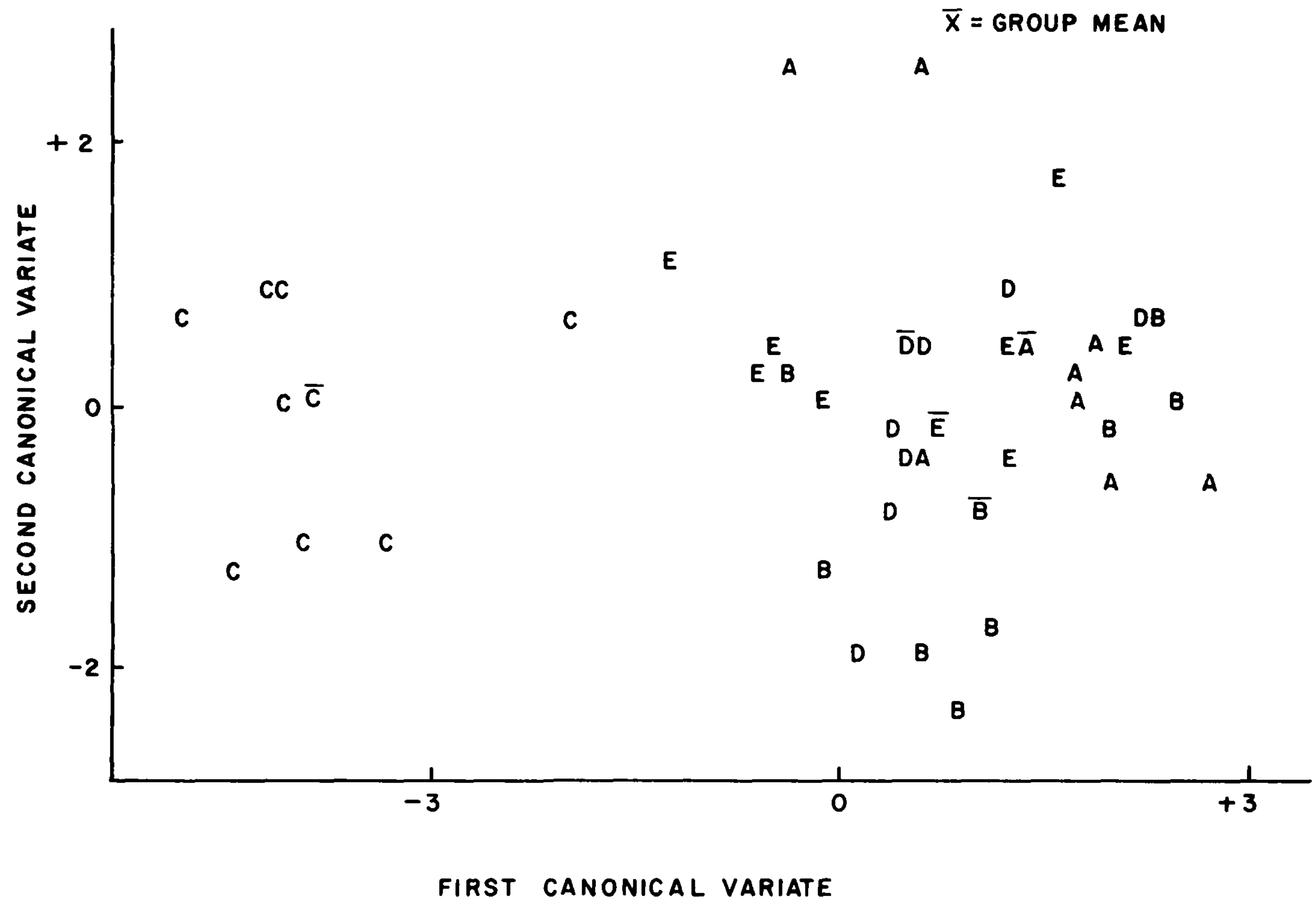


Figure 30. Plot of the Two Canonical Variates for Experiment XI

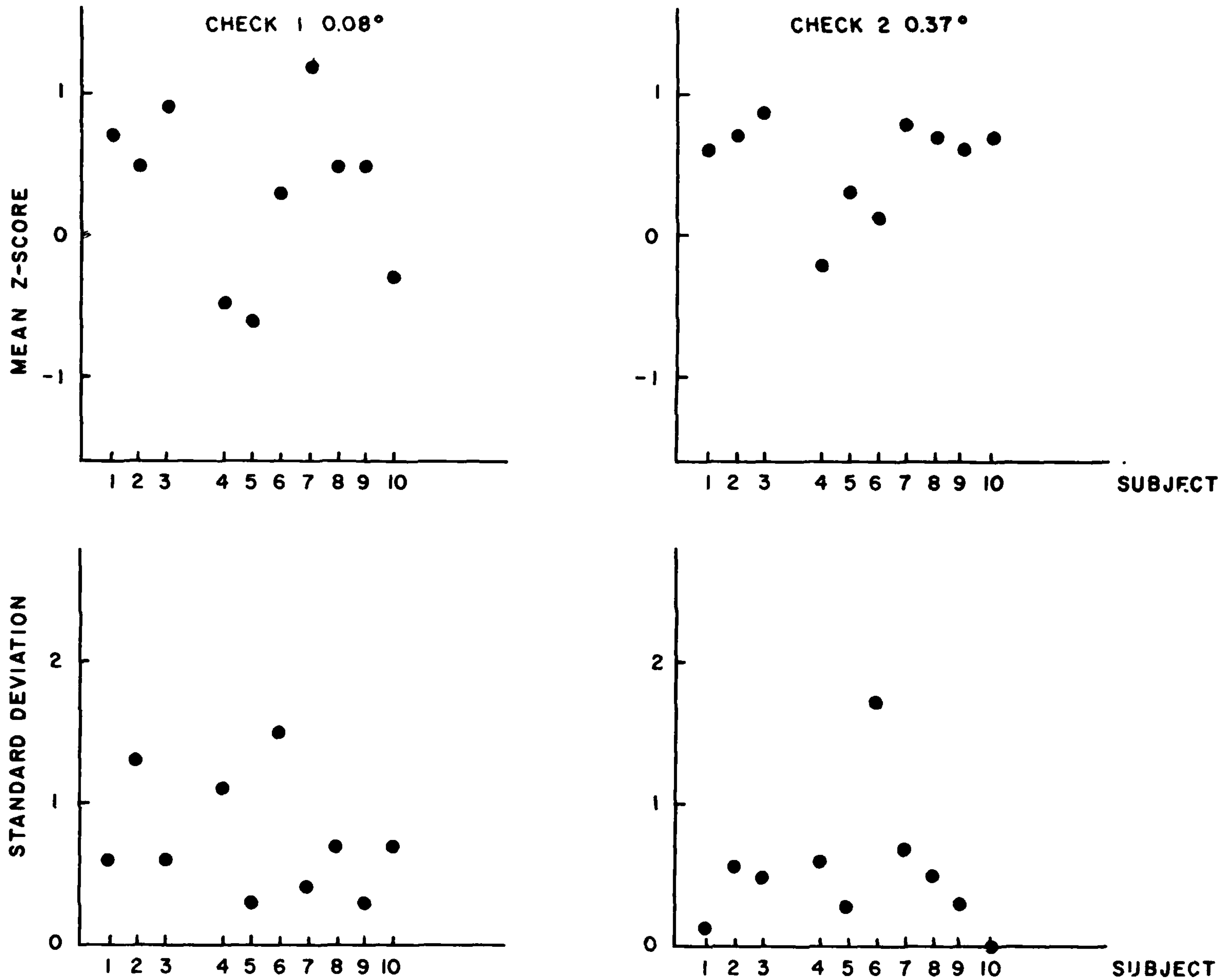


Figure 31. Z-Score Means and Standard Deviations for Windscreen A

TABLE 29. STANDARD DEVIATIONS FOR THE TWO CHECKERBOARD STUDIES

Phase	Check	Windscreen A	B	C	D	E
I	1	.226	.093	.190	.150	.263
	2	.133	.143	.134	.218	.101
II	1	.637	.321	.704	.538	.260
	2	.361	.270	.951	.501	.339
Combined	1	.502	.304	.295	.150	.361
Analysis	2	.226	.483	.358	.360	.433

In addition to the variability problem, some Phase II subjects used a very restricted range of responses. Even though some subjects used a large number of distinct responses, they tended to restrict their responses to a small portion of the 0 to 1 scale. Other subjects used the entire scale, but repeatedly used a few scale values. The use of these repeated values was not consistent across replications. These effects are displayed in Table 30.

TABLE 30. THE RANGE AND NUMBER OF DISTINCT VALUES JUDGED BY SUBJECTS IN PHASE I AND PHASE II ACROSS ALL WINDSCREENS

Subject	Checkerboard 1 0.18°			Checkerboard 2 0.37°		
	Range	Difference	Number Values	Range	Difference	Number Values
1	.15-.95	.80	10	.15-.95	.80	10
2	.20-1.00	.80	11	.25-1.00	.75	9
3	.20-1.00	.80	15	.15-.95	.80	11
4	.30-.90	.60	9	.00-.60	.60	8
5	.50-.90	.40	6	.50-.90	.40	7
6	.80-1.00	.20	8	.70-1.00	.30	7
7	.75-.95	.20	7	.15-.53	.38	10
8	.70-.97	.27	13	.65-.97	.32	14
9	.25-.97	.72	10	.40-.97	.57	11
10	.00-1.00	1.00	9	.00-.90	.90	5

In addition to the above problems, Phase II subjects seemed to use the judgment scales differently than Phase I subjects did. Windscreen B was judged to be worse and windscreen E was judged to be better by subjects in

Phase II, while the reverse was true in Phase I. There was virtually no overlap and, hence, two very different distributions resulted. Naturally, these aberrant responses did not lead to effective windscreen discriminability.

3. EXPERIMENT XII: LINEAR MOTION BAR

In this experiment, the basic Phase I study was repeated with 10 subjects; three were naive and the others had participated in some of the previous studies. One additional target condition, a very thin bar, was added to the Phase I conditions. This was done because the thin bar target tended to "break up" when it was viewed through severe distortions. Table 31 shows the conditions for the linear motion bar target.

TABLE 31. LINEAR MOTION BAR TARGET CONDITIONS

Bar	Size in Degrees Visual Angle
1	0.07
2	0.15
3	0.30

The results of the analysis showed that bar 3 was the "best" single variable. The classification chart, shown in Table 32, reveals that rather poor discrimination was obtained. Only 48% of the windscreens were correctly classified.

TABLE 32. CLASSIFICATION CHART FOR JUDGMENTS BASED ON BAR 3

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	0	3	0	7	0
B	3	2	0	1	4
C	0	0	10	0	0
D	3	1	0	6	0
E	1	3	0	0	6

Figure 32 shows the Z-score means when bar 3 was the target.

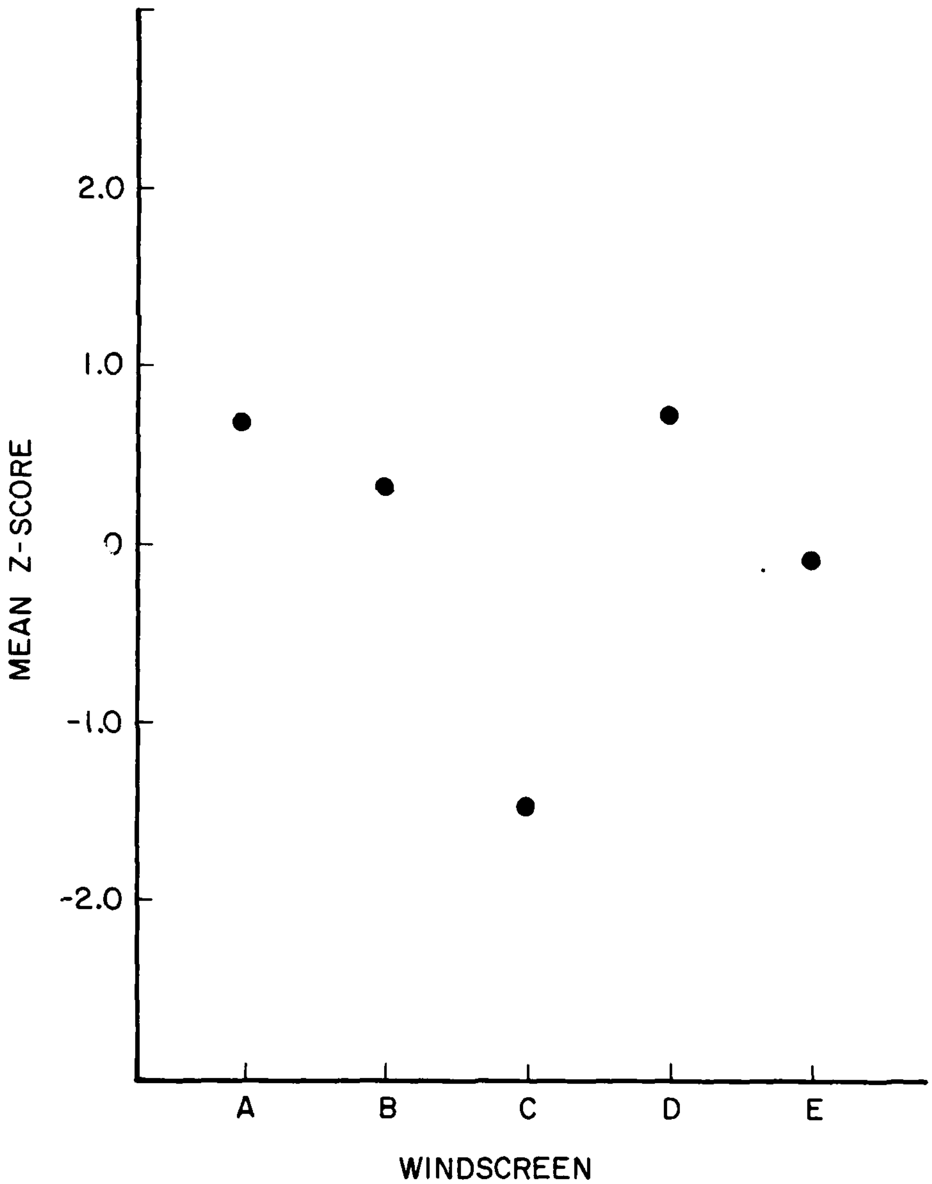


Figure 32. Z-Score Means (Judgments) with Bar 3 as the Target

Clearly, the results of this study are not in agreement with the Phase I study. In the early work, the best target was the 0.15° bar, not the 0.30° bar that was found to be "best" in Phase II. Many of the subjects used in the Phase II work were also those who participated in the Phase II checkerboard study. The problems with the linear motion bar experiment are probably similar to those encountered in the checkerboard experiment. Some insight into the problem is afforded by comparing the standard deviations from the Phase I and Phase II studies. Standard deviations for six subjects were higher for the Phase II work; only two were lower. Apparently subject variability is, in part, responsible for the failure to obtain agreement in the two studies.

4. CONCLUSIONS, PHASE II

Both Phase II studies show poorer discrimination than was obtained in the Phase I pilot work. The principal reason for this has been attributed to the higher variability in subjects' responses. One might be tempted to assert that Phase I subjects performed well because they had considerable experience by the end of the pilot work. It should be recalled, however, that the grid study was the first experiment run and that these subjects performed well from the onset of the work. Further, subjects did not reduce their variability as work in a given experiment progressed. The experimental conditions were blocked across five replicates and the standard deviations tended to remain constant across replicates. Subjects did not "learn" to make more accurate judgments. An analysis of variance with replicates as a factor showed that subjects did not increase their discriminability among windscreens or a function of practice. (Experiment XI, Checkerboard, $F = 2.49$; $df = 4, 36$, n.s.; Experiment XII, Linear Motion Bar, $F = .41$; $df = 4, 36$, n.s.)

For any visual inspection technique to have merit, the subjective rating must be reliable. Based on the discrepancies found in subject performance between Phase I and Phase II, it would seem that judgment reliability would be facilitated by using experienced observers and training them to use the whole scale of available values.

SECTION IV
PHOTOGRAPHIC STUDIES

1. APPROACH

These studies were undertaken to see if ratings of windscreen distortion would be similar when photographs of targets taken through windscreens were used as opposed to viewing the targets directly. Such a procedure offers the advantage that a permanent record would exist for each windscreen and judgments could be made by observers without the windscreen being present.

2. SUBJECTS

Ten subjects were used in this study. Each had participated in at least one other windscreen study and two had served in all of the studies.

3. APPARATUS

Four of the "best" targets were chosen and photographs of each were taken through the five windscreens. In addition, the four targets were also projected as in the Phase I and Phase II studies for real-time direct-view evaluation of windscreen distortion. The four static targets chosen and their physical characteristics are presented in Table 33. (Naturally, dynamic targets could not be used for the photographic study.)

TABLE 33. TARGET CHARACTERISTICS FOR THE PHOTOGRAPHIC STUDIES

Target	Parameters
1	0.275° grid; black lines on white background
2	0.15° horizontal grating
3	0.37° checkerboard
4	0.275° grid; white lines on black background

4. PROCEDURE

Each subject served in both the photographic and direct-view studies. Half of the subjects judged the photographs first and half began with the windscreens. Target order was randomized for a given subject, but once an

order was established, a subject completed all trials for a given target before changing targets. As before, windscreens (or photographs) were presented in a random order across the blocks of targets. Judgments were made using the magnitude estimation procedure. All viewing was binocular.

5. RESULTS

The general results of this work are summarized in Table 34, which presents the percentage of correct classifications for each target type. Each target was analyzed by discriminant analysis and then the data from the four targets for each study were combined to analytically determine the "best" target condition for the two rating methods.

TABLE 34. PERCENT CORRECT WINDSCREEN CLASSIFICATION FOR REAL-TIME JUDGMENTS AND JUDGMENTS BASED ON PHOTOGRAPHS

Target	Windscreen Rated	Photograph Rated
1. Black line grid	42	38
2. Grating	52*	40
3. Checkerboard	44	56*
4. White line grid	54	48
Linear combination		
2 and 1	54	
3 and 2		64

Overall, discrimination tended to be slightly better when actual windscreens were rated rather than photographs. Photograph ratings were superior when a linear combination of a checkerboard and grid were computed. A plot of the first two canonical variates for judgments based on photographs is [illegible] in Figure 33. It is apparent that the only reliable discrimination [illegible] between windscreen group C and the other groups. Table 35 shows the [illegible]ion chart derived from the second step of the discriminant analysis [illegible] 3 and 2 were analytically combined. Only 64% of the windscreens [illegible] classified.

[illegible]ingle target based on the discriminant analysis comparisons

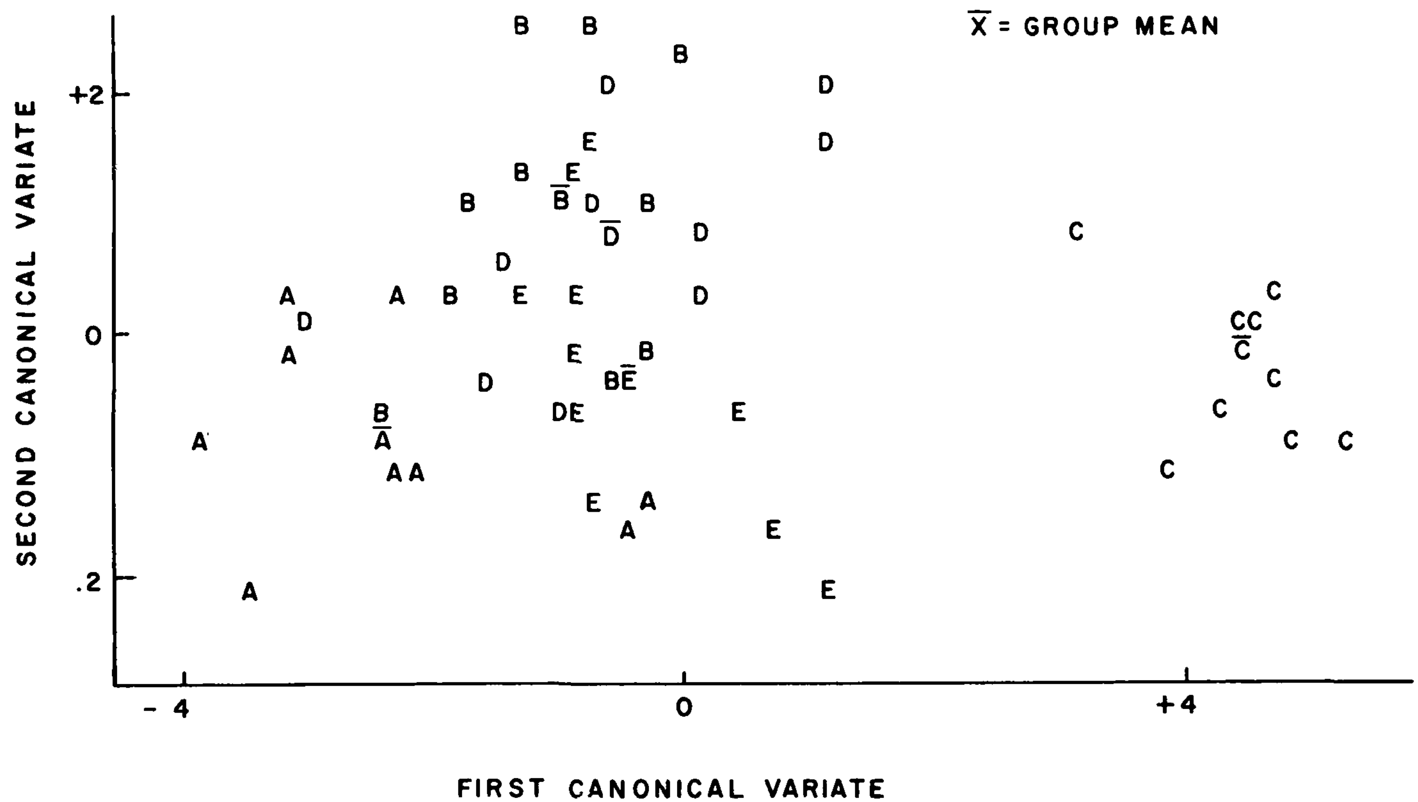


Figure 33. Plot of the First Two Canonical Variates for Judgments Based on Photographs

TABLE 35. CLASSIFICATION CHART FOR JUDGMENTS BASED ON PHOTOGRAPHS

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	7	1	0	0	2
B	2	5	0	2	1
C	0	0	10	0	0
D	2	1	0	5	2
E	1	2	0	2	5

When subjects judged the same targets through the windscreens, a linear combination of a grid and a grating (targets 2 and 1) afforded only 54% correct classification. These data are shown in Table 36.

TABLE 36. CLASSIFICATION CHART FOR JUDGMENTS BASED ON WINDSCREENS

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	5	0	0	4	1
B	1	5	1	0	3
C	0	0	10	0	0
D	4	1	0	3	2
E	0	3	0	3	4

Figure 34 shows a plot of the two canonical variates derived from the discriminant analysis. The data are similar to that obtained in the study on photographs. Group C is the only one that is clearly separated from the others.

6. DISCUSSION

The results of the work on photographs and windscreens can be compared with the outcome of previous studies. For instance, when the data of the grating experiment (VII) are compared, some interesting similarities emerge. That study employed both monocular and binocular viewing conditions for a 0.15° grating. In a sense, the present studies may mirror these two viewing conditions. Both studies employed direct-view assessment with binocular viewing. In the current work, judgments made on the basis of photographs

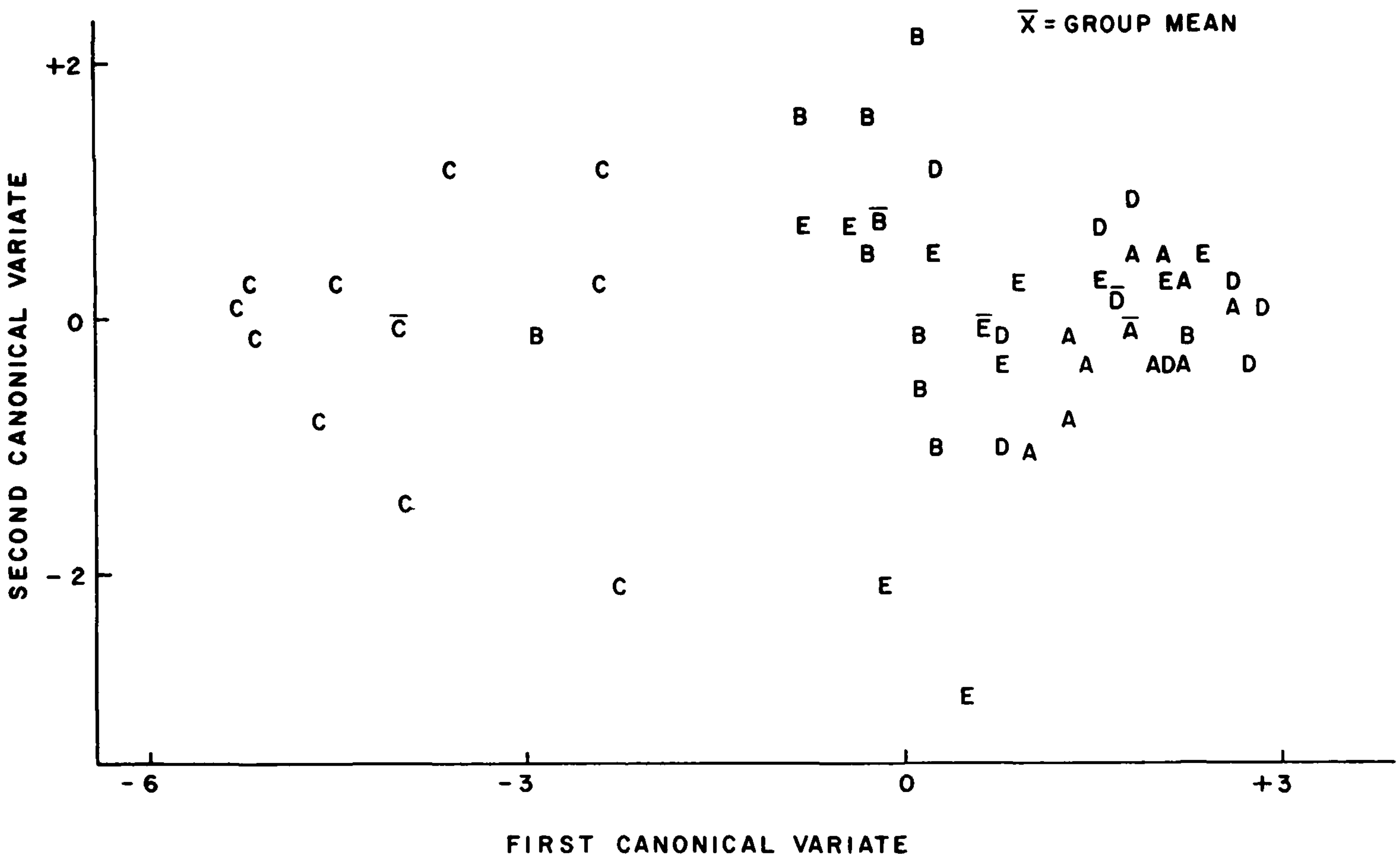


Figure 34. Plot of the Two Canonical Variates for Judgments Based on Windscreens

may resemble monocular viewing in that they cannot profit from information dependent on retinal disparity. A comparison of the data from the two studies is shown in Figure 35. Histograms of mean Z-score ratings permit comparisons of the monocular and binocular viewing conditions and show striking similarities. The mean ratings in the monocular grating study (Experiment VII) closely resemble those determined for the photographs. When subjects binocularly viewed the grating target through the actual windscreen, the pattern of ratings was similar for Experiment VII and the present study.

It appears that the subjects' judgments in the present series of studies (described in Section IV) are similar to those obtained in earlier work. The poor discrimination seemed to be due to large variability in subjects' responses. When analysis was performed on the response pattern across subjects, targets, and windscreens, effects similar to those reported for the checkerboard experiments (Experiment XI) were found. Subjects often utilized narrow ranges of the 0-1 response scale and were not systematic in their use of the scale. No subjects could be identified who failed altogether to exhibit some discrimination, so there was no basis to exclude any data in hopes of improving the analysis.

7. CONCLUSIONS

On the basis of the analyses presented, it must be concluded that no evidence exists as to the profitability of rating photographs in lieu of rating windscreen distortion while viewing targets in real-time. The evidence is not entirely clear because response variability could have masked differences in the two procedures.

The data do suggest that the use of photographs may be more limited than real-time viewing because no cues useful for retinal disparity exist. The data of Experiments VII and VIII show that discrimination is best when information is utilized from both monocular and binocular viewing. Viewing of photographs is also more limiting in that no information can be derived from motion parallax. Motion parallax can be made available easily in the direct-view inspection situation.

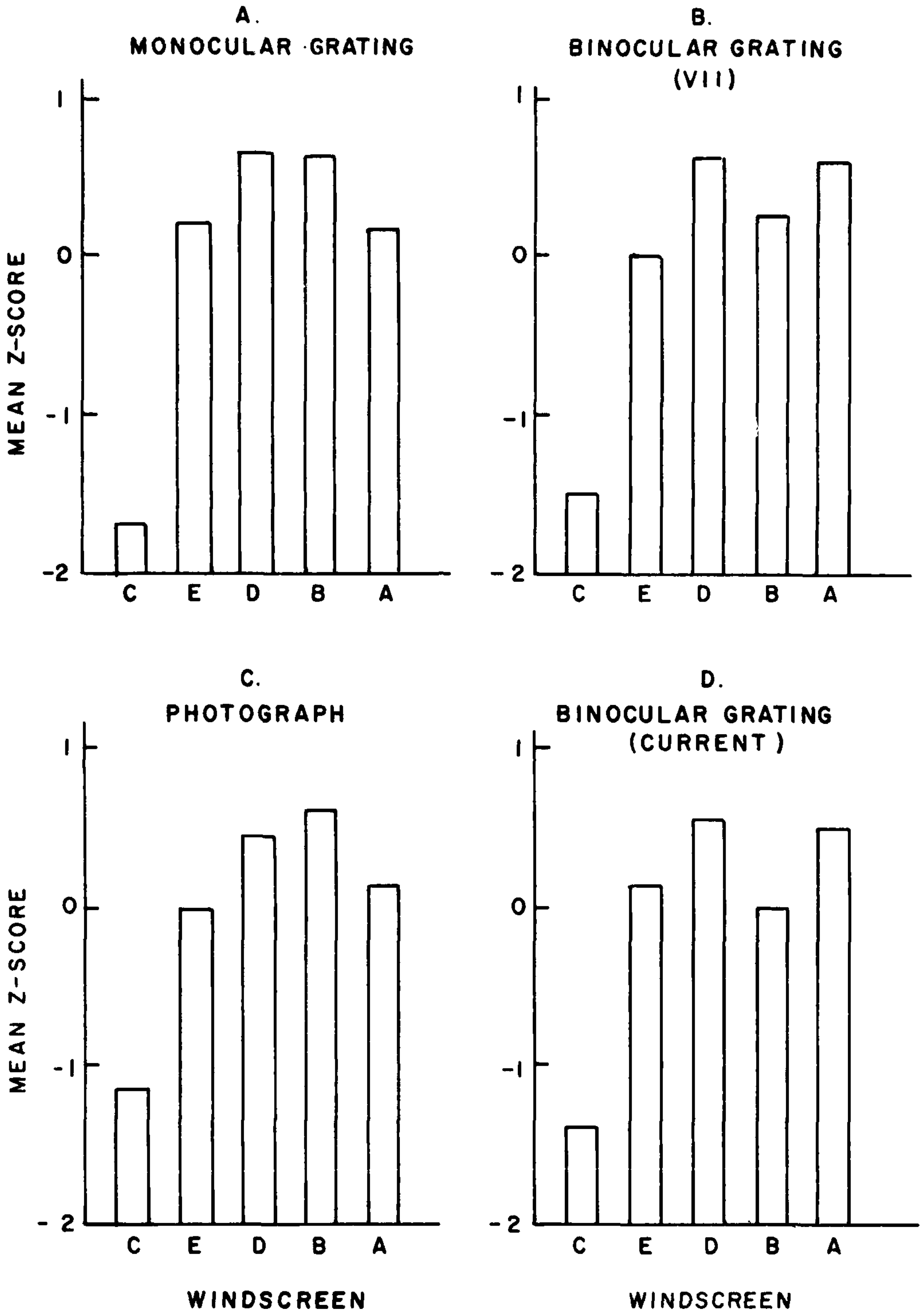


Figure 35. A Comparison of Mean Z-Scores for Binocular and Monocular Viewing in Experiment VII (A and B) and the Present Study (C and D)

SECTION V

SCALING VIA COMPUTER-GENERATED IMAGERY

1. APPROACH

The previous studies all utilized a subjective scaling approach, i.e., magnitude estimation. While the data clearly show that such scaling can be successful (if subjects follow directions and utilize the entire scale), the ability to compare magnitude estimation with one of the more classical psychophysical procedures was desired.

In this final study, subjects were required to match windscreen distortion with one of a series of computer-generated patterns. In essence, subjects were forced to compare the windscreen-induced distortion with a scale of distortion which was defined and anchored to physical attributes. This approach, of course, presents the risk that the defined scale may be inappropriate or may fail to represent fully the nature of windscreen distortion. However, it has an advantage in that it requires subjects to make judgments which can be referenced to known physical dimensions. In addition, it was thought that subjects might be more likely to utilize the entire range of matching stimuli since they could view all the stimuli while making their judgments.

2. SUBJECTS

Ten subjects participated in this study. Some of them had served in earlier windscreen studies.

3. APPARATUS

A series of "images" was plotted using SYMVU*, a computer package that allows the user to specify the X, Y, and Z coordinates of an array. Using an external program, a series of sinewaves that were exponentially damped in the X dimension and quadratically damped in the Y dimension was generated. The displays were rotated to an azimuth of 30° and were viewed at an altitude of 45° to enhance the wave-like character of the images.

---

*SYMVU is available from the Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, Harvard University.

Figure 36 shows the nine images that were used. The letters E, L, Q, M, N, G, F, K, and J are arbitrary codes assigned to the computer-generated images (CGI). Each image was varied in two dimensions. Exponential damping was varied using values of 0.1 to 1.0 as the damping factor in the equation:

$$Z = Y(2-Y)A\exp(-DX)\cos(FX^2)$$

where A = Amplitude of the center wave

D = Damping factor

F = Frequency (fixed at 10).

Values of Y varied from 0 to 2 and values of X from 0 to 10 for each figure. Amplitude of the center sinewave was varied from 0.12 to 0.37 according to the scaling assigned the SYMVU program. The actual plots were 14 by 16.5 cm.

The wave equation was designed to represent two common properties of windscreen distortion. First, the proportion of the windscreen viewing aperture that contained distortion was simulated by the damping factor. Damping controlled the spatial extent of the waves or ripples in the CGI figures. Secondly, it was desired that the CGI figure reflect the fact that the wave-like patterns of distortion in the windscreens seemed to vary in height or amplitude. Therefore, the SYMVU plotting routine was used to vary the height of the CGI plots to stimulate this effect. By covarying these two physical dimensions, it was anticipated that the CGI stimuli might lead to better windscreen discrimination than the simpler univariate magnitude judgments.

4. PROCEDURE

First, subjects rank-ordered the set of nine CGIs. They were told merely to order the set of stimuli along whatever dimension seemed to make sense to them (see Appendix A). Once subjects had ordered the stimulus set, the CGI figures were always available as a set in the order which each subject had established. Subjects were told to look at the target through a windscreen and then look through the ordered CGI set and pick out the one that best matched the type of distortion they observed. Subjects were urged to look back through the windscreen as often as necessary while selecting the matching CGI.

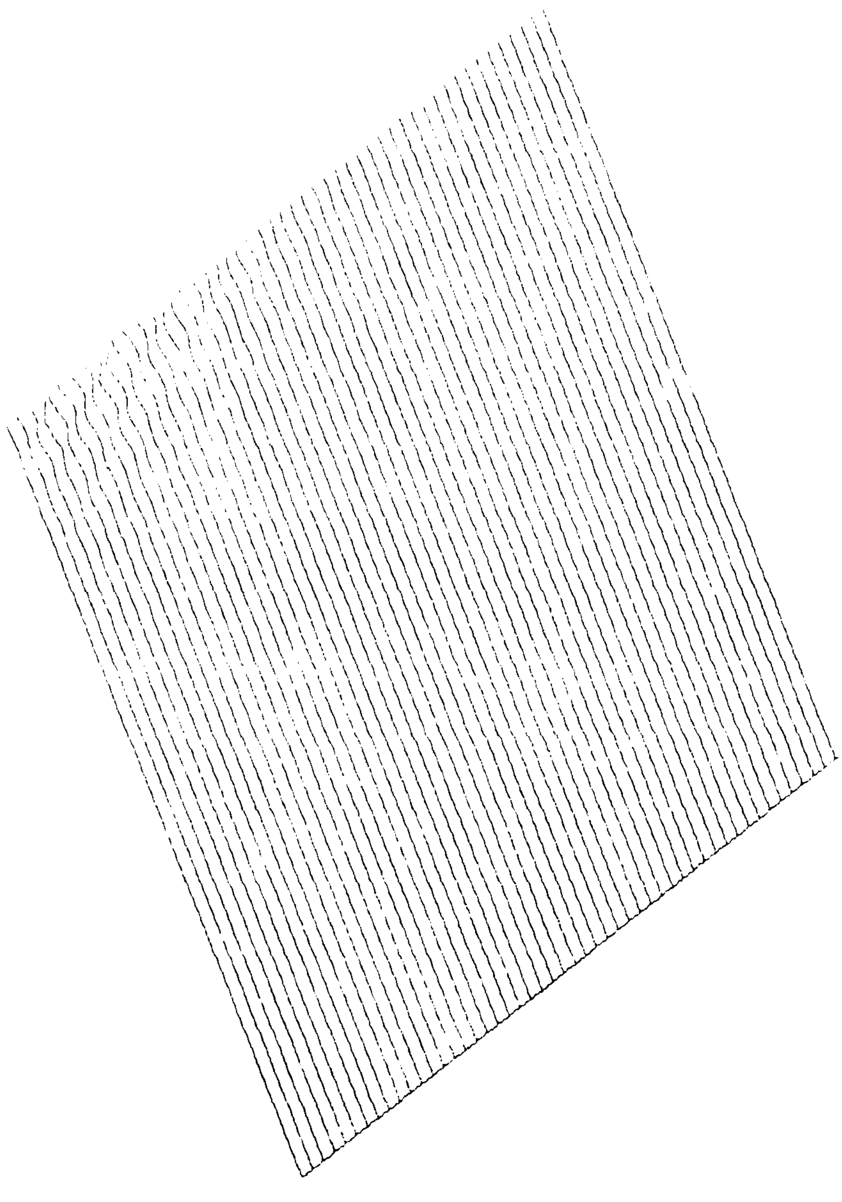


Figure 36E. CGI with Amplitude 0.12 and Damping 1.0

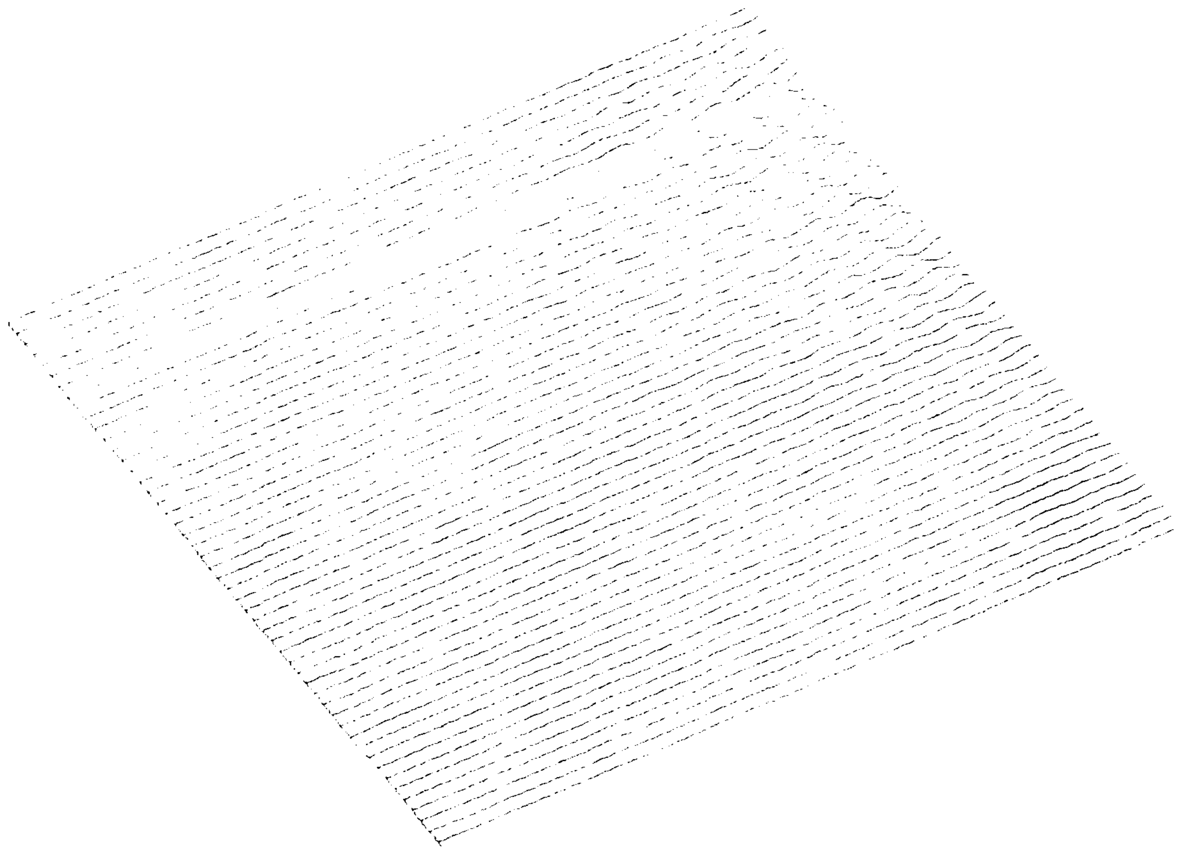


Figure 36L. CGI with Amplitude 0.12 and Damping 0.5

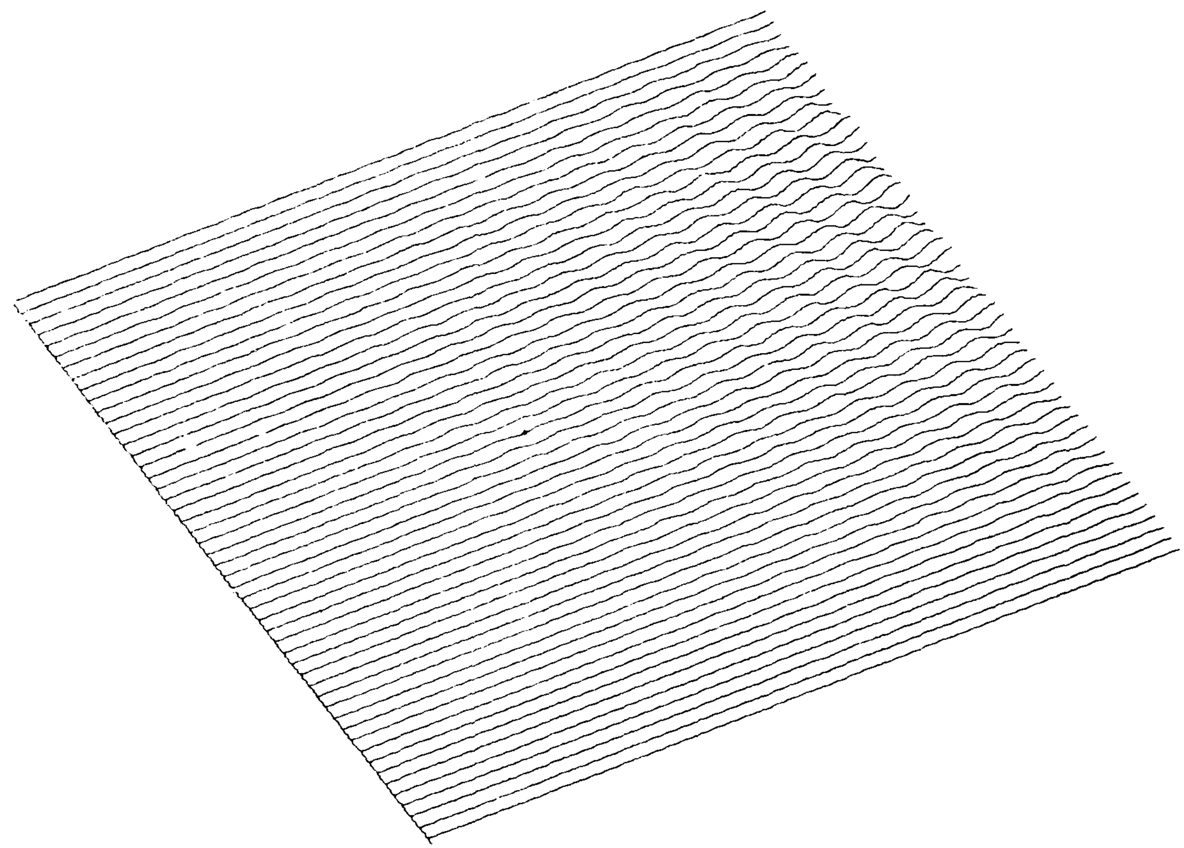


Figure 36Q. CGI with Amplitude 0.12 and Damping 0.25

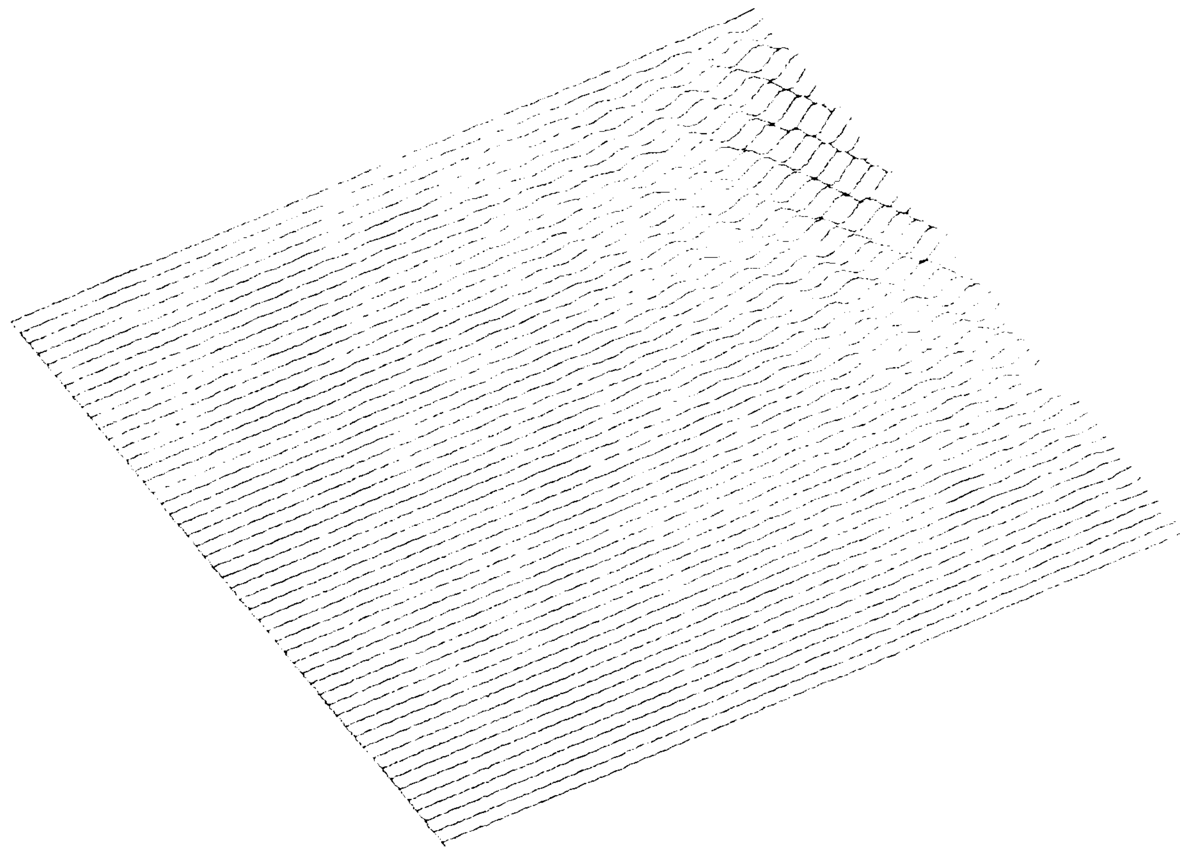


Figure 36M. CGI with Amplitude 0.25 and Damping 0.5

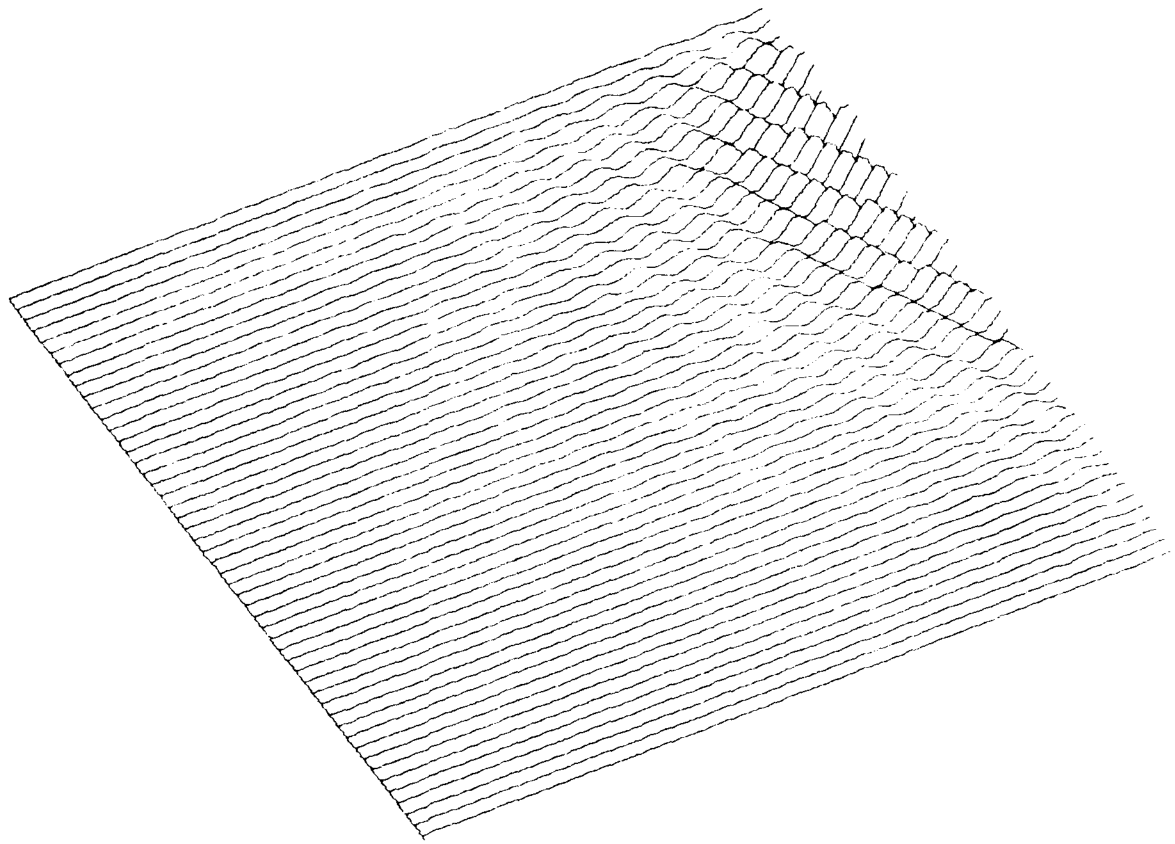


Figure 36N. CGI with Amplitude 0.37 and Damping 0.5

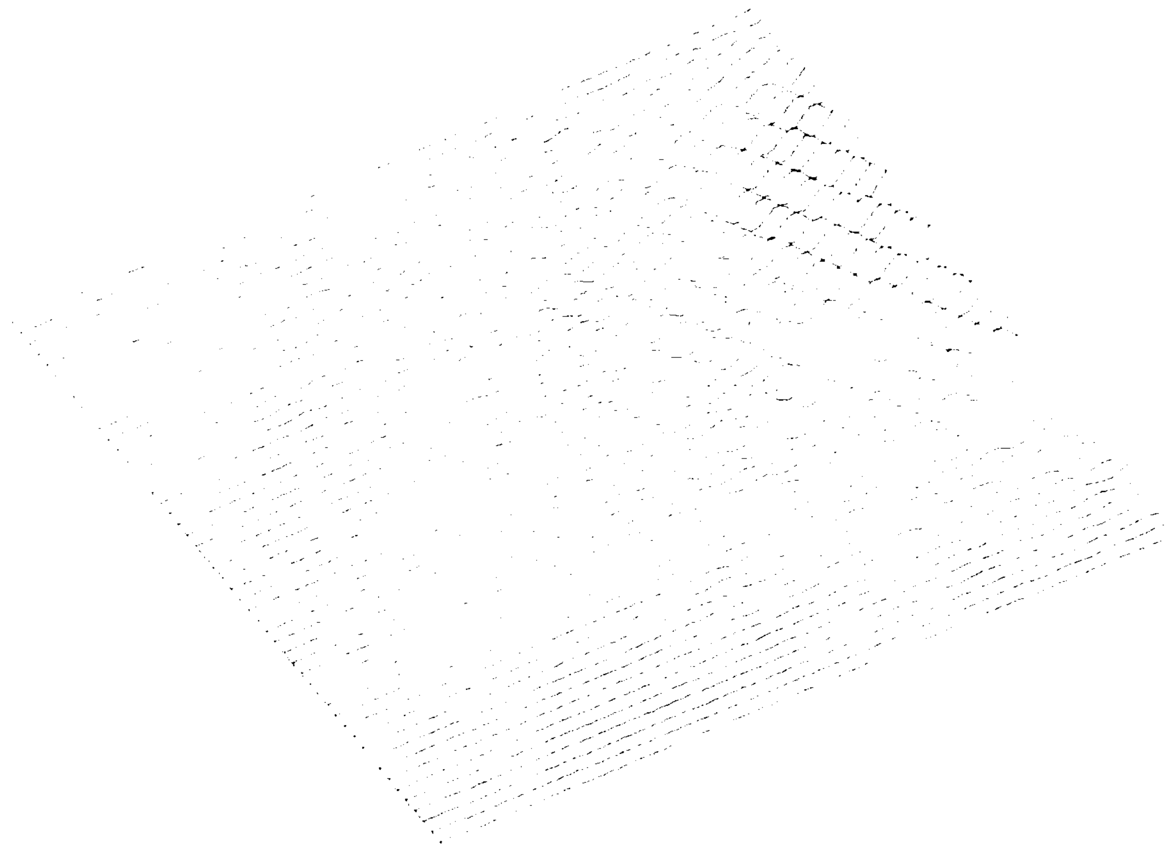


Figure 36G. CGI with Amplitude 0.25 and Damping 0.25

AD-A079 369 SYSTEMS RESEARCH LABS INC DAYTON OH ELECTRONIC WARFA--ETC F/G 1/3
DEVELOPMENT OF A VISUAL INSPECTION TECHNIQUE (OPTICAL ASSESSMEN--ETC(U)
OCT 79 F E WARD , A J DEFRANCES F33615-77-C-0535

UNCLASSIFIED AMRL-TR-79-67 NL

2 OF 2
AD
A079369

END
DATE
FILMED
2-80
DDC

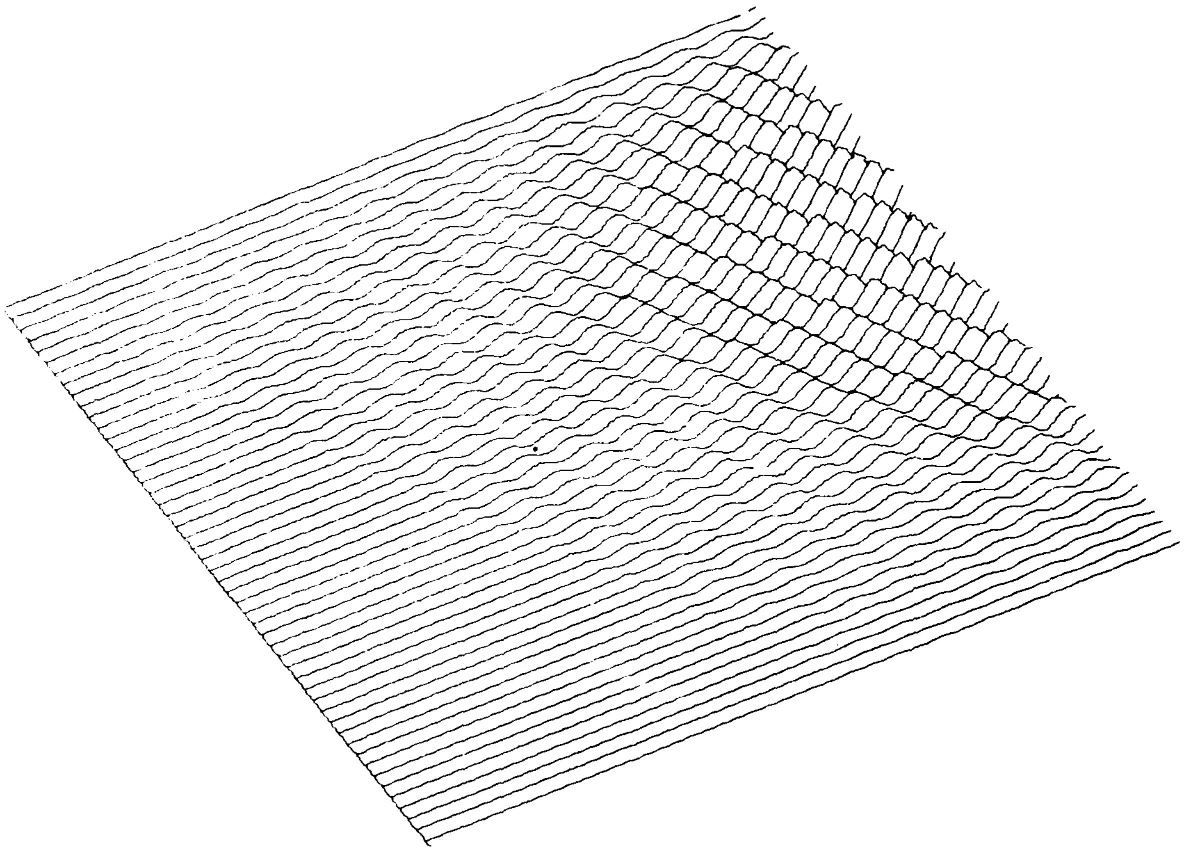


Figure 36F. CGI with Amplitude 0.37 and Damping 0.25

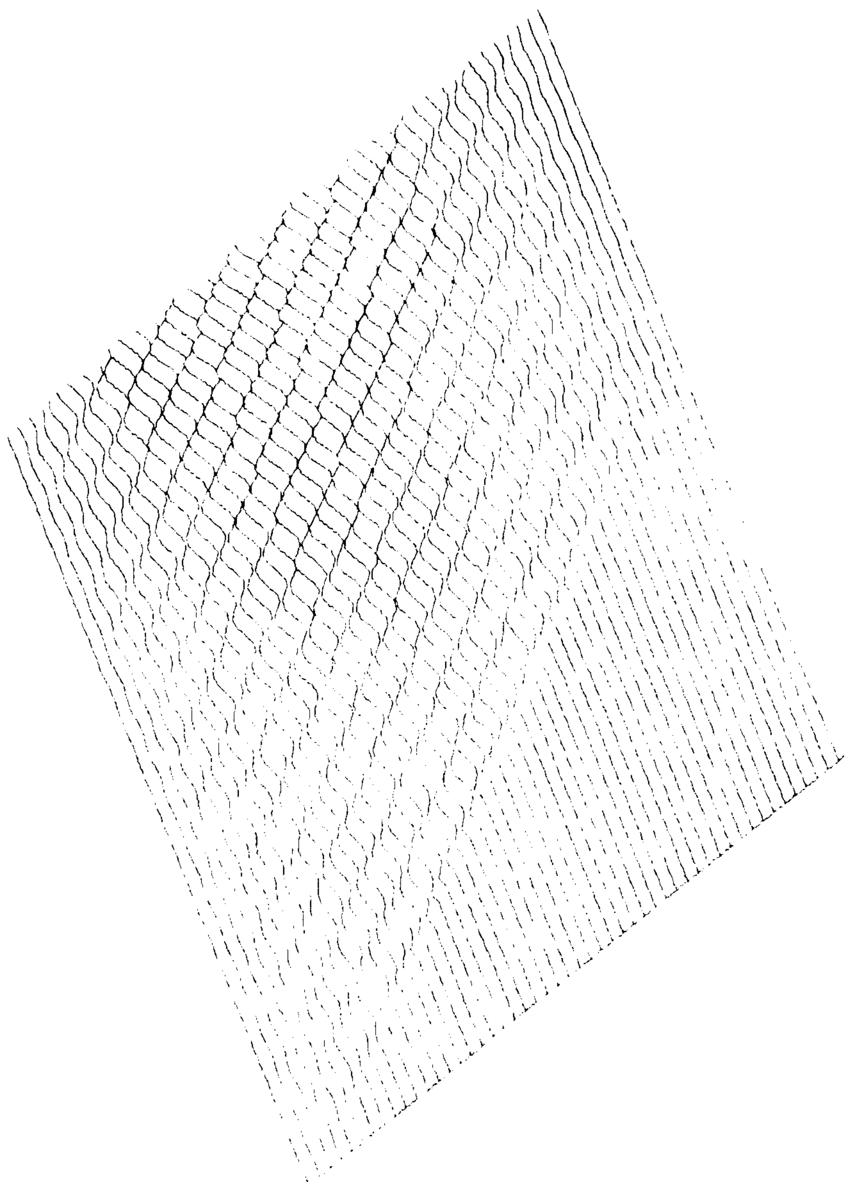


Figure 36K. CGI with Amplitude 0.25 and Damping 0.1

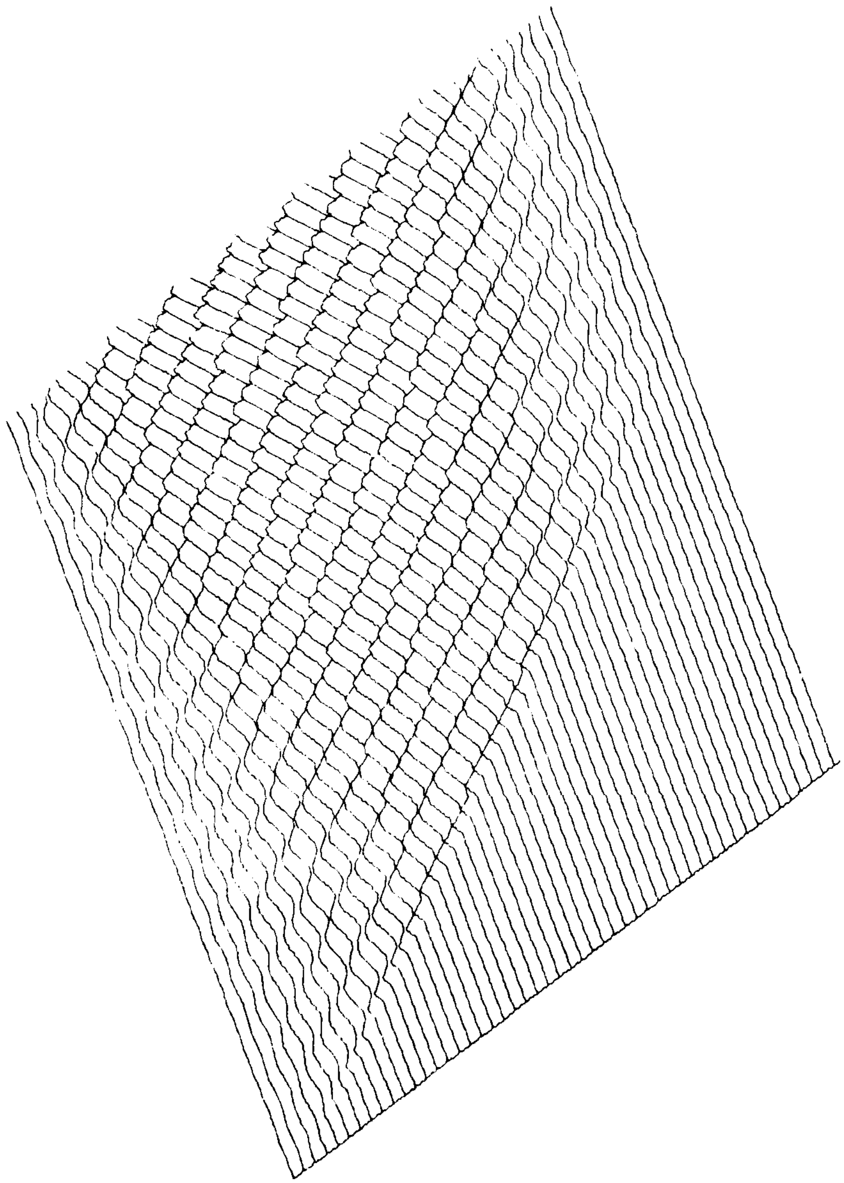


Figure 36J. CGI with Amplitude 0.37 and Damping 0.1

Two targets were used in the CGI study, a grid and a dynamic bar. The black grid lines were spaced about 0.275° apart on a white background (grid 3 of Experiment I). The dynamic target was a horizontal bar, 0.30° thick (bar 3, Experiment XII) that moved vertically at 2°/sec.

5. DISCRIMINANT ANALYSIS

Each subject's judgments were transformed into a ranking by assigning the matched CGI stimulus the rank that corresponds with that subject's pre-trial rank ordering of the CGI stimuli. The ranks were then entered into the discriminant analysis program and analyzed in the same fashion as the magnitude estimation data. No other transformations were performed on the ranked data. It should be noted that this approach was not strictly appropriate because the ranks represent a discrete scale rather than the continuous scale that is proper for the analysis.

6. RESULTS

The discriminant analysis program extracted two canonical variates based on the targets. Figure 37 shows a plot of the canonical variates in which it is clear that no group discrimination is apparent. The bar target afforded the most discrimination for a single variable, 51%. The linear combination of both targets produced 56% correct classification, the details of which are presented in Table 37.

TABLE 37. CLASSIFICATION CHART FOR JUDGMENTS BASED ON THE BAR AND GRID

Windscreen Tested	Windscreen Classed As				
	A	B	C	D	E
A	4	3	0	2	0
B	1	5	1	1	1
C	0	0	6	1	2
D	4	0	0	4	1
E	0	1	1	1	6

The first canonical variate accounted for most of the discrimination. The canonical correlations for the first and second canonical variates were 0.815 and 0.390, respectively.

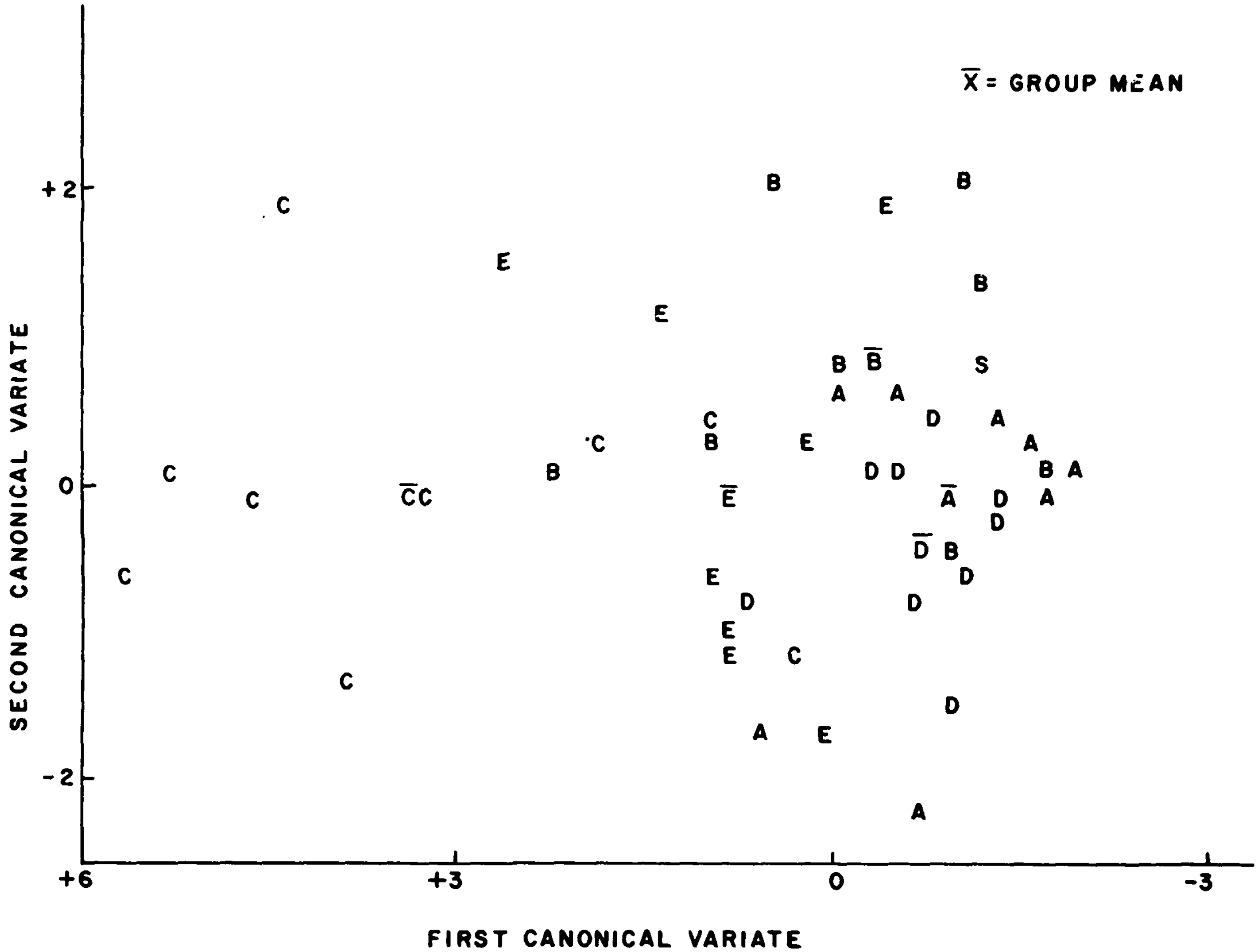


Figure 37. Plot of the Two Canonical Variates from the CGI Study

The above analysis is based on data from nine subjects. One subject's data had to be eliminated because he failed to exercise any discrimination for the different windscreens. The other nine subjects did appear to discriminate and were able to do the task easily according to their introspective reports.

Subjects had very good agreement on the pre-trial rank-order of the CGI stimuli. Table 38 shows the letter code of the stimulus with the number of times it was ranked in the order indicated. Rankings which were atypical did not show any particular pattern and were attributed to random errors rather than confusion concerning interpretation of the CGI patterns.

TABLE 38. RANK ORDER FOR THE CGI FIGURES

	1	2	3	4	5	6	7	8	9
Stimulus Order	J	K	F	G	N	M	Q	L	E
Number of Times Ranked	9	9	9	7	7	7	7	10	10

Since the discrimination analysis approach was not totally appropriate for ranked data and classification was poor, the CGI data were analyzed by two alternative procedures. First, the frequency of each rank assigned to each windscreen was plotted for the two targets, the static grid and the dynamic bar. Figure 38A-E and Figure 39A-E show these frequency data for the two target types. The figures show that, in general, windscreen A was judged best for both target types. Windscreen D tended to be ranked as the second best, although the data is not as clear for the bar target. Windscreen C was always ranked worst and produced the most variability in the rankings.

The data were also analyzed by considering the physical dimensions of damping and height that were varied to construct the CGI figures. Damping and height are plotted in Figure 40A-E. At each appropriate intersection of these two variables is indicated the number of times that particular CGI figure was chosen to represent the given windscreen. The first frequency value is for the grid and the second value corresponds to the dynamic bar

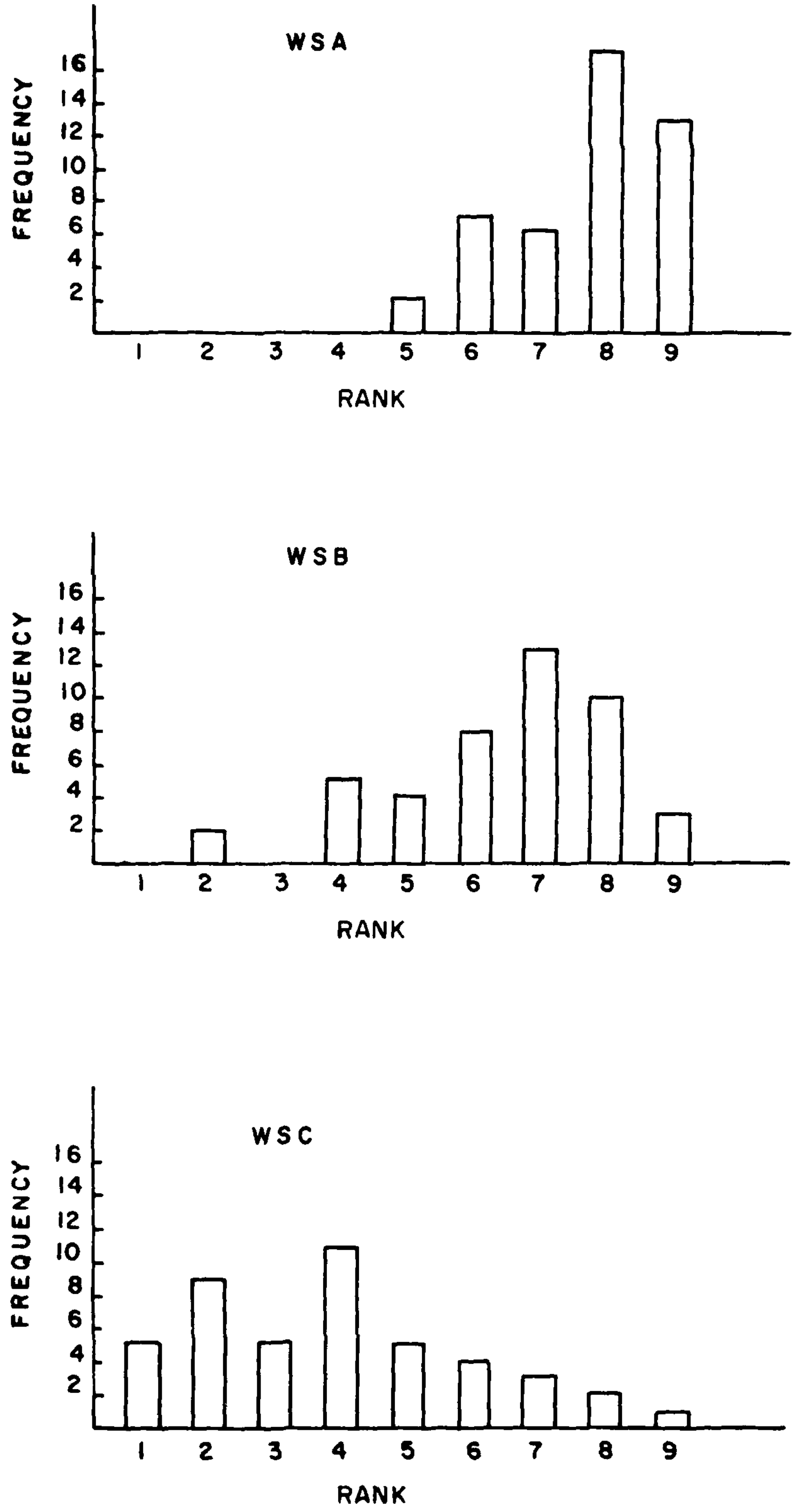


Figure 38A-C. Frequency of Ranks for Windscreens A-C for the Grid Target

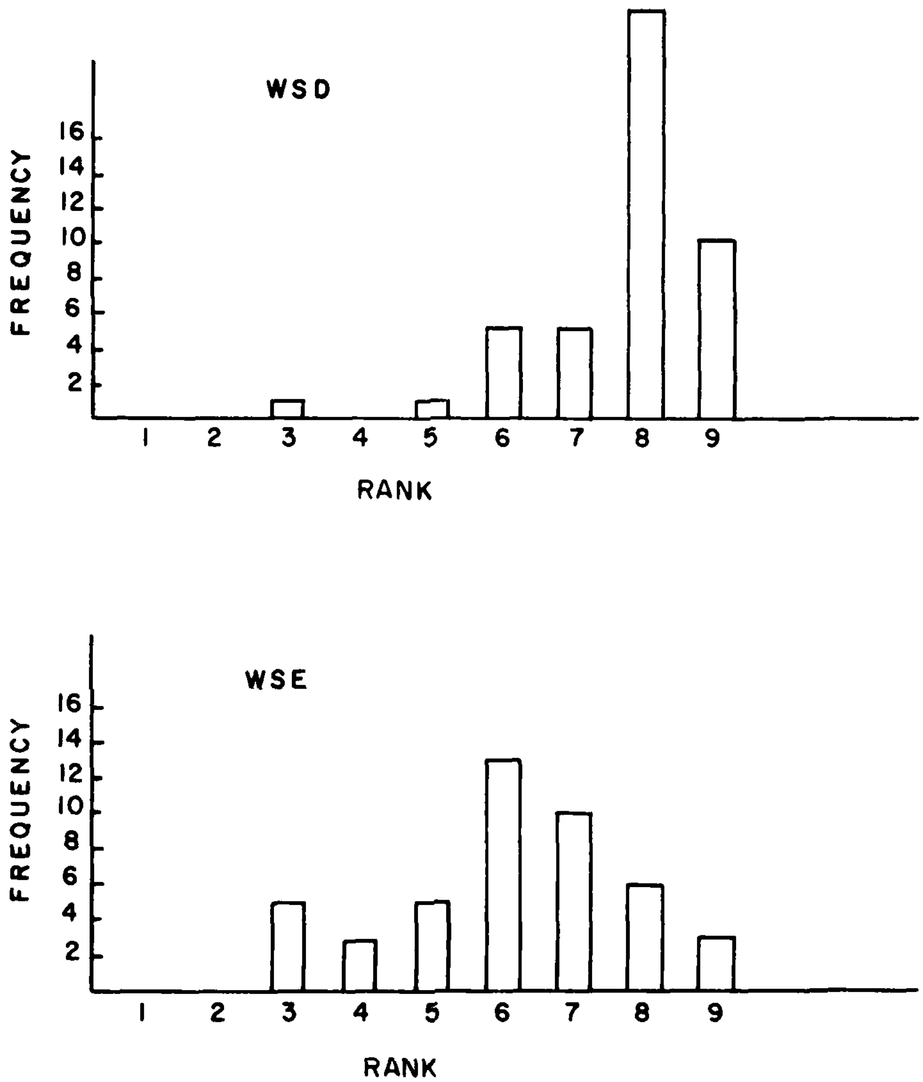


Figure 38D, E. Frequency of Ranks for Windscreens D,E for the Grid Target

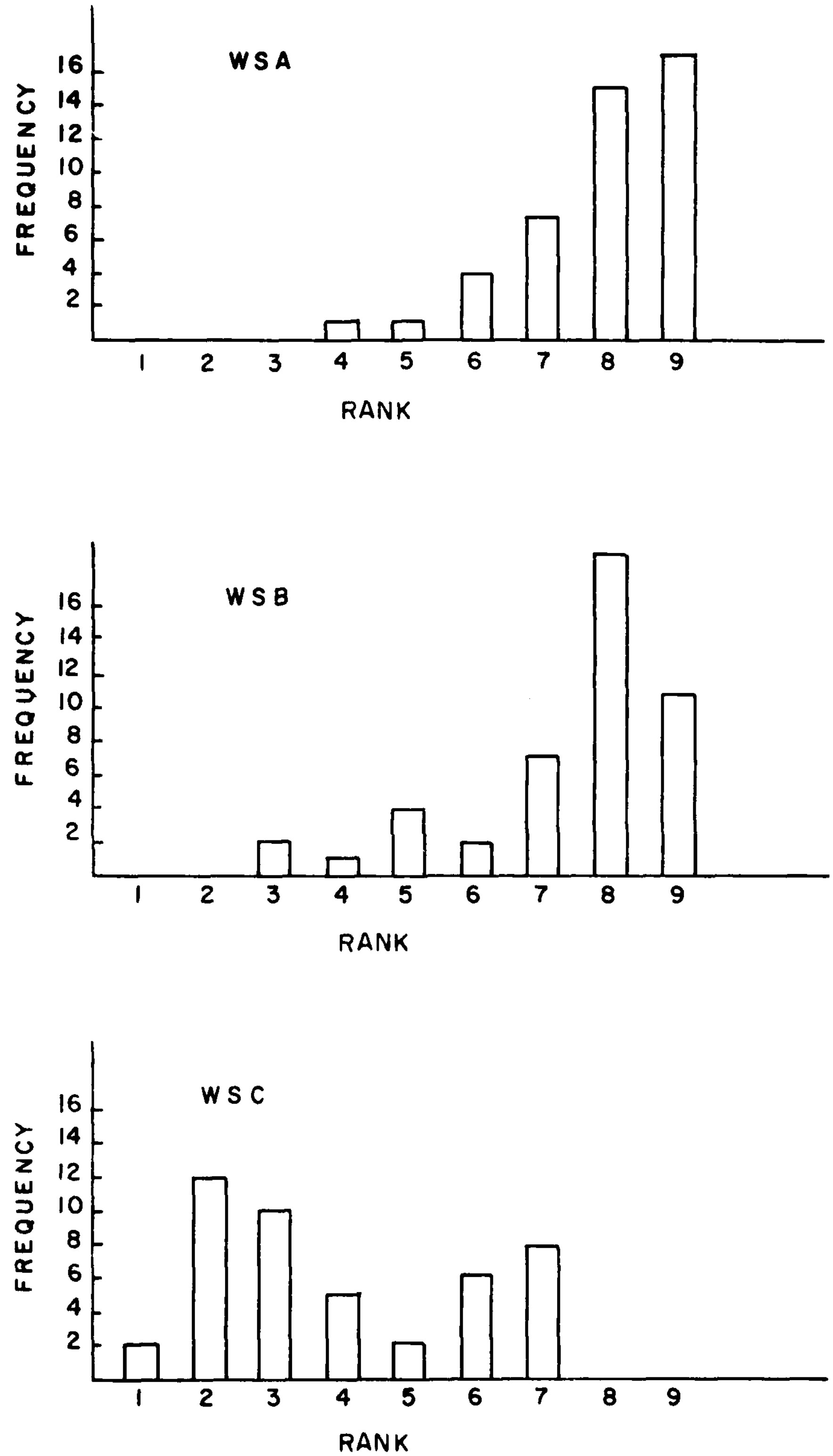


Figure 39A-C. Frequency of Ranks for Windscreens A-C for the Dynamic Bar Target

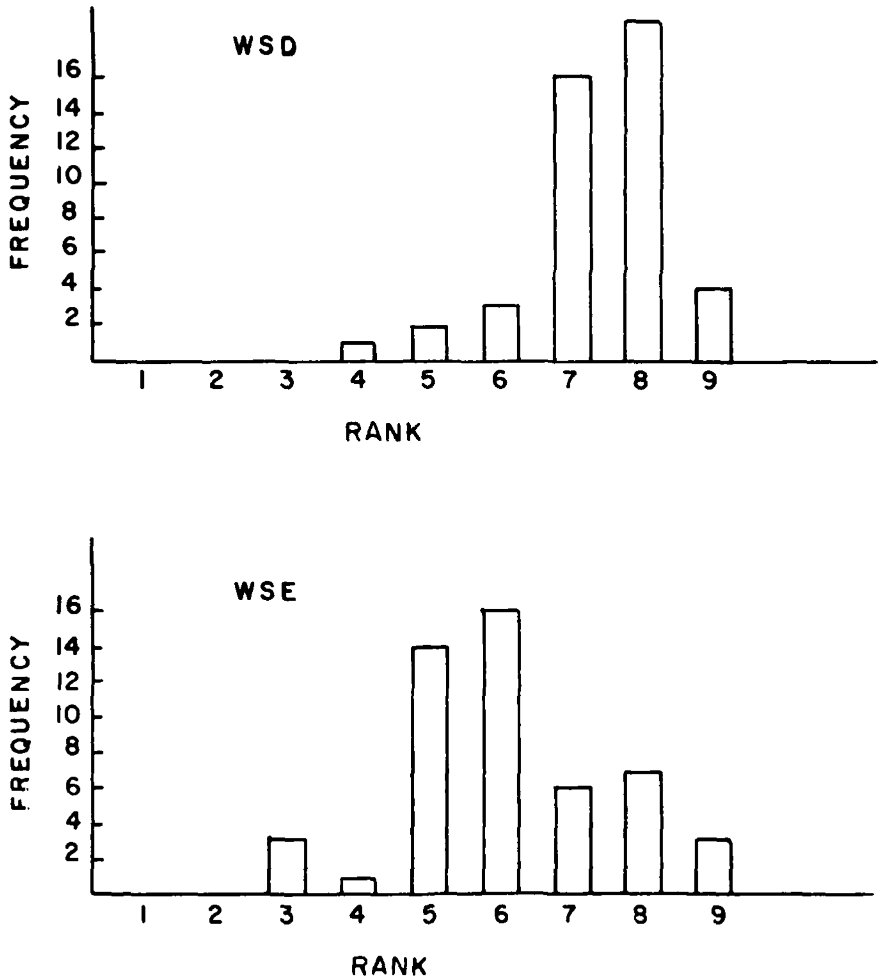


Figure 39D, E. Frequency of Ranks for Windscreens D,E for the Dynamic Bar Targets

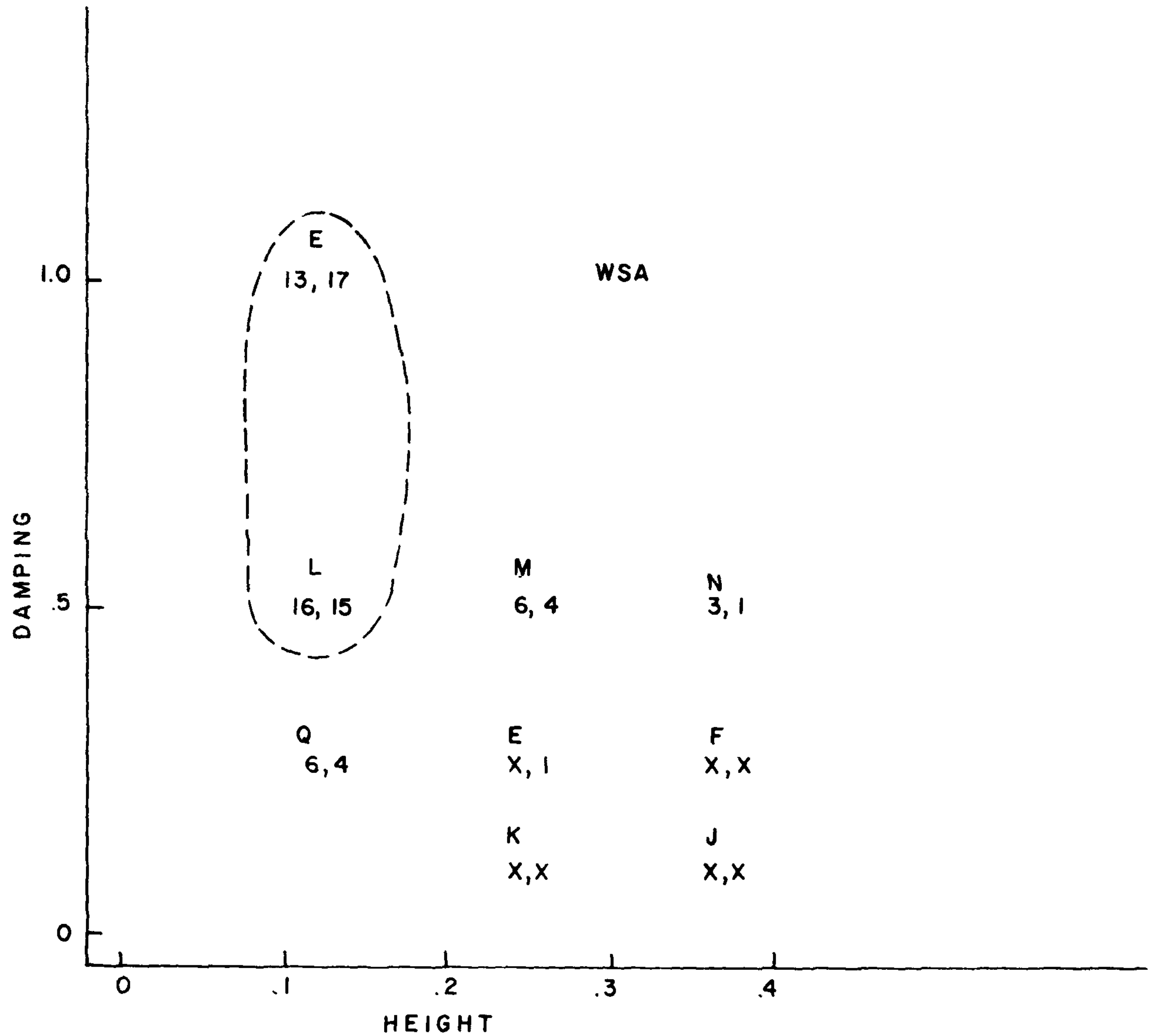


Figure 40A. CGI Frequencies Matched to Windscreen A

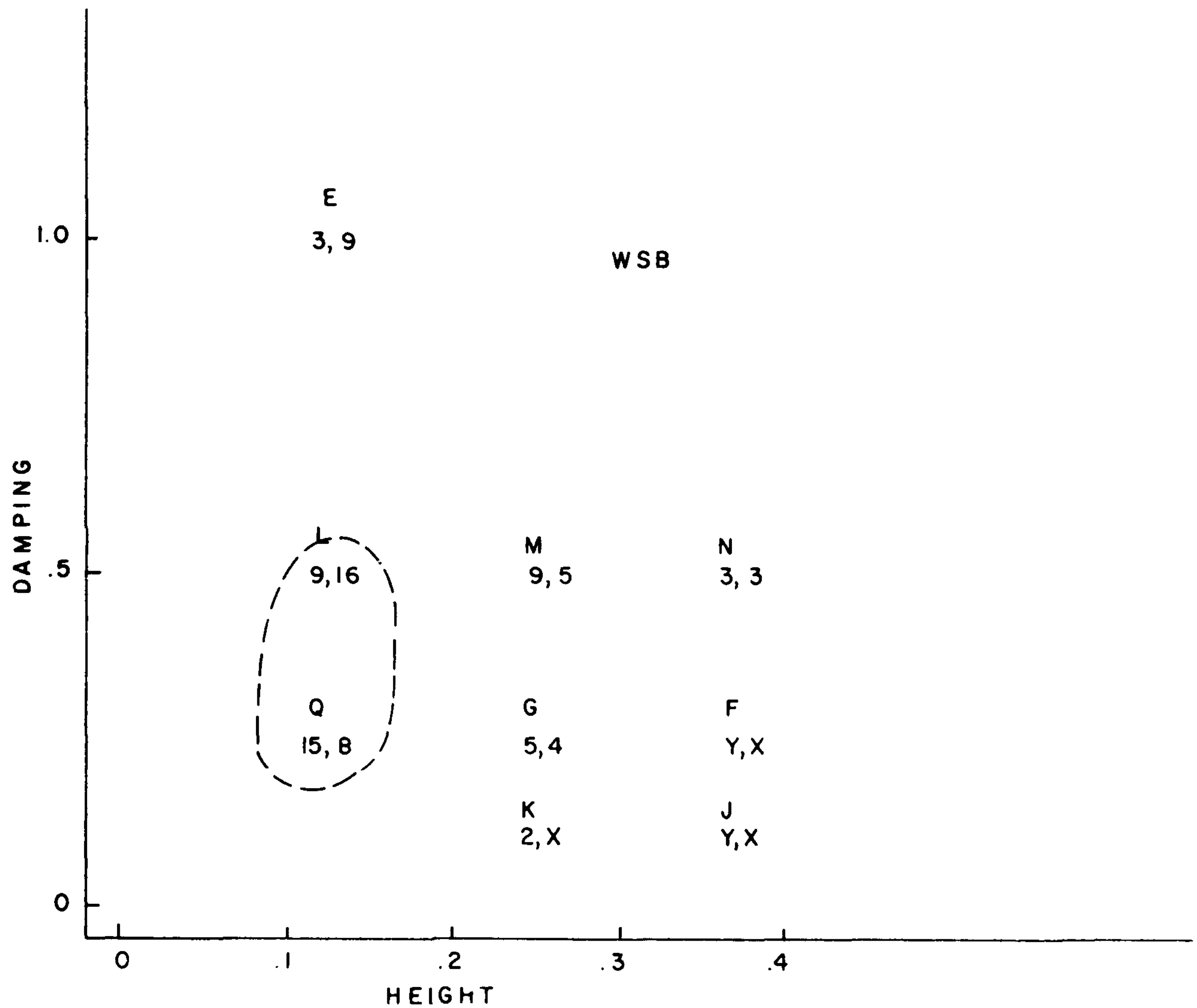


Figure 40B. CGI Frequencies Matched to Windscreen B

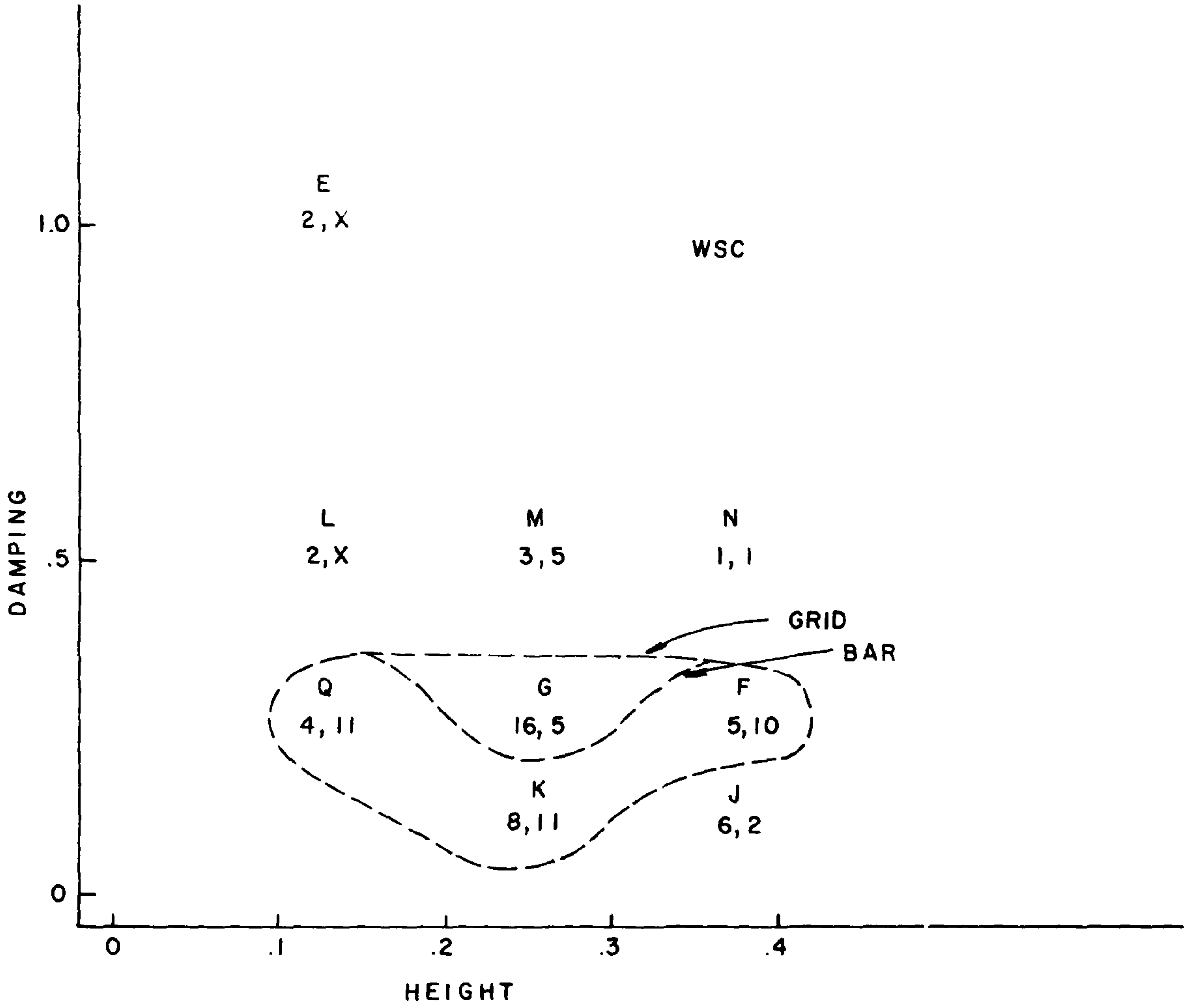


Figure 40C. CGI Frequencies Matched to Windscreen C

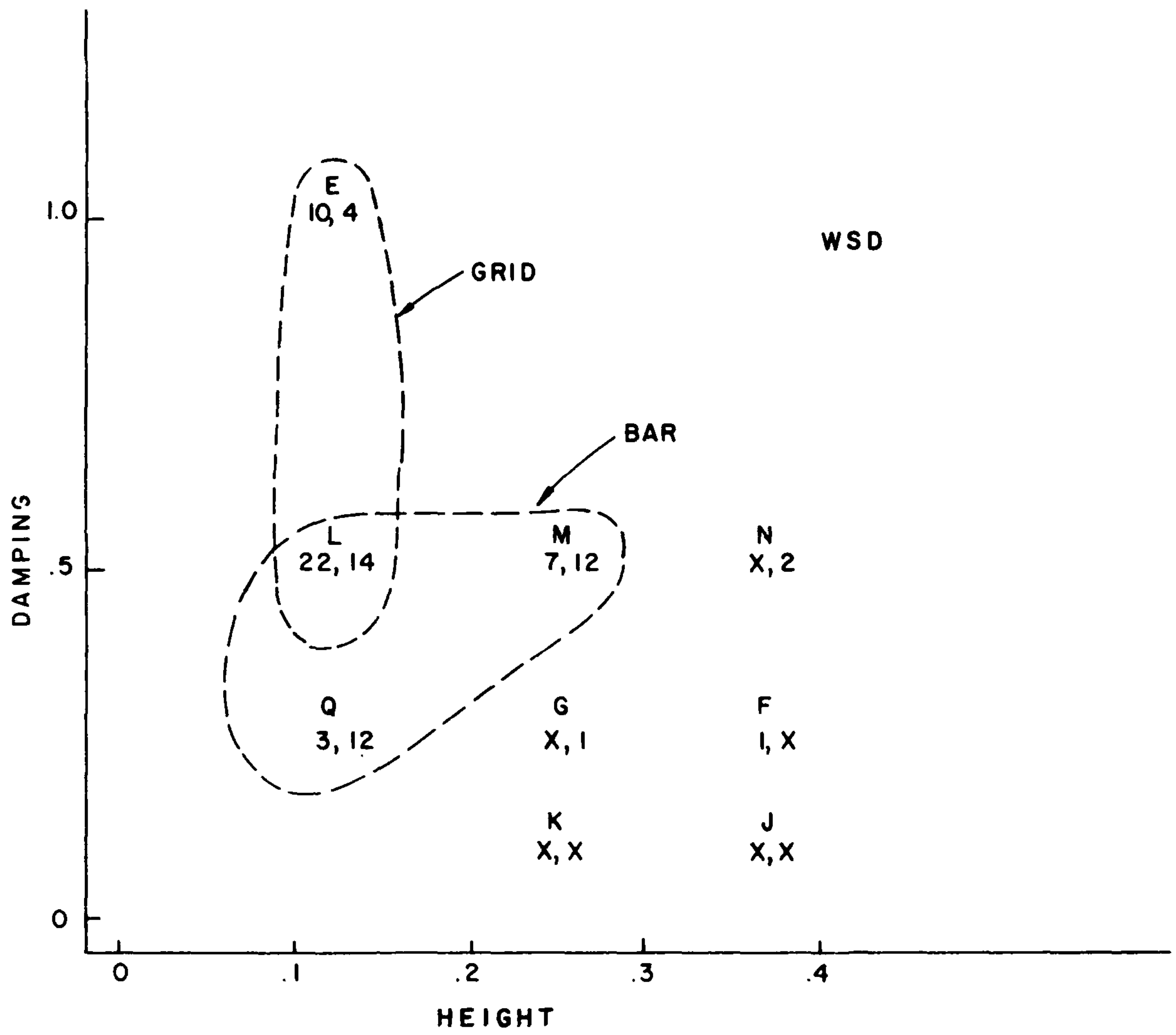


Figure 40D. CGI Frequencies Matched to Windscreen D

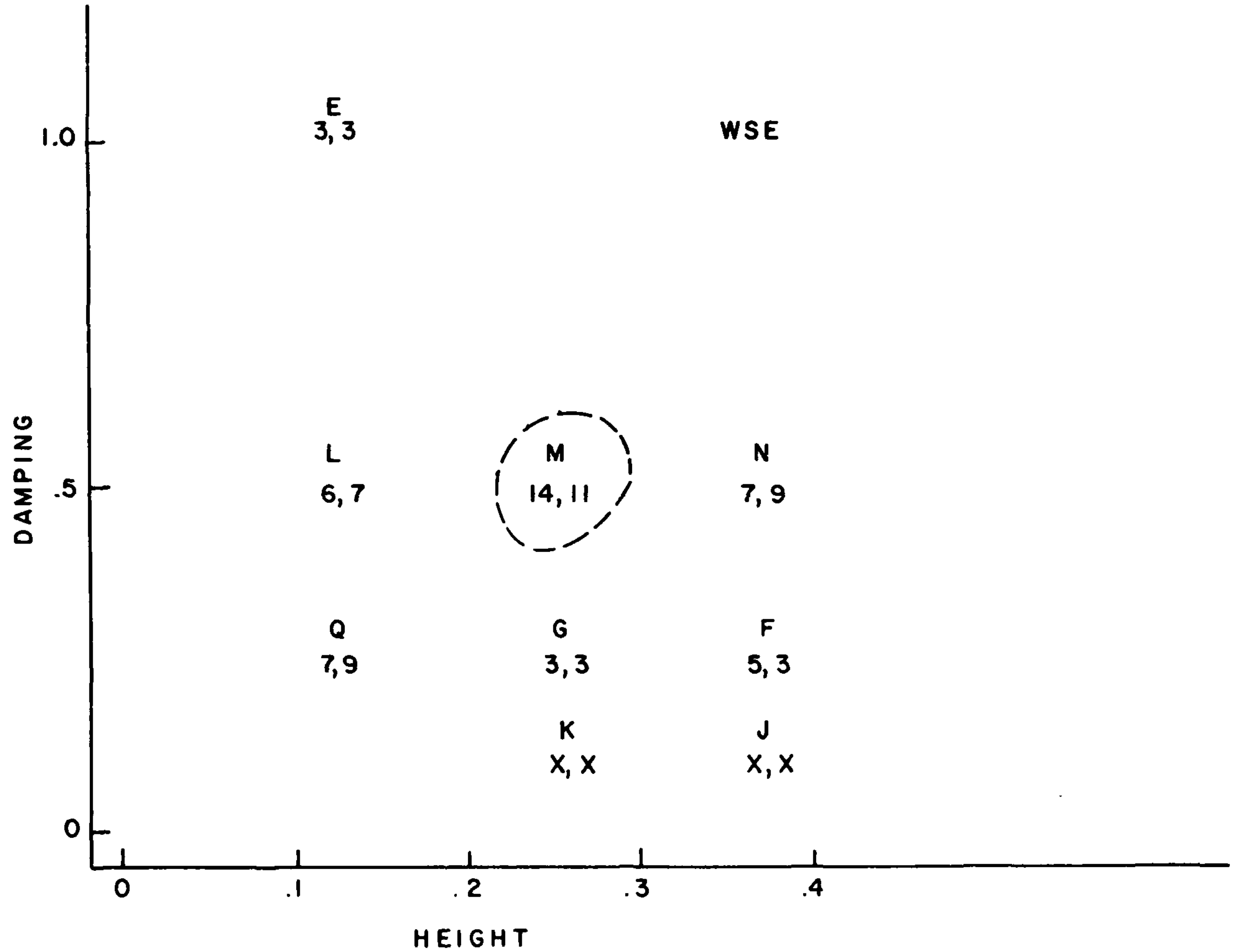


Figure 40E. CGI Frequencies Matched to Windscreen E

target. The letter indicates the relevant CGI designation. The location in the damping-height plane that is most often chosen to represent each windscreen has the highest frequency rates and is outlined with dashed lines.

It is apparent that each of the five windscreens occupies a particular subset of the damping-height space. In the case of windscreen D, there is considerable overlap of these subspaces, although the actual pattern of overlap varies slightly with the target type. In general, the bar target proved to afford slightly better discrimination. When the subspace areas for all five windscreens are plotted together, it is apparent that windscreen D is most often confused with B and E. Figure 41 shows this in schematic plot of the windscreen locations based on judgments made with the dynamic bar target.

7. DISCUSSION

The physical attributes, height and damping, form a two-dimensional space that is useful for categorizing windscreen-induced distortion. Windscreens A and C are easily separated although windscreen D is confused with B and E. Were windscreen D not included, discrimination would actually be quite good. Apparently windscreen D embodies distortions (ripples) that are difficult to match to CGI stimuli of the type used in this study.

A potential difficulty with our CGI matching procedure is that the two-dimensional space defined by the CGI stimuli may not represent the middle of the range very well. This middle area seems to be precisely where the greatest confusion occurs. The procedure could be made more sensitive by providing CGI stimuli at intermediate points in the damping-height space.

Of course, there is an alternative explanation for the confusion in the middle ranges of stimulus attributes. Perhaps the CGI scale is adequate and subjects are simply unable to make fine discriminations. This explanation was ruled out because our subjects demonstrated ease and precision in their pre-trial ranking of the stimuli. If the CGI stimuli represented attribute step sizes that were below subjects' threshold, the ranks could not have been so consistent.

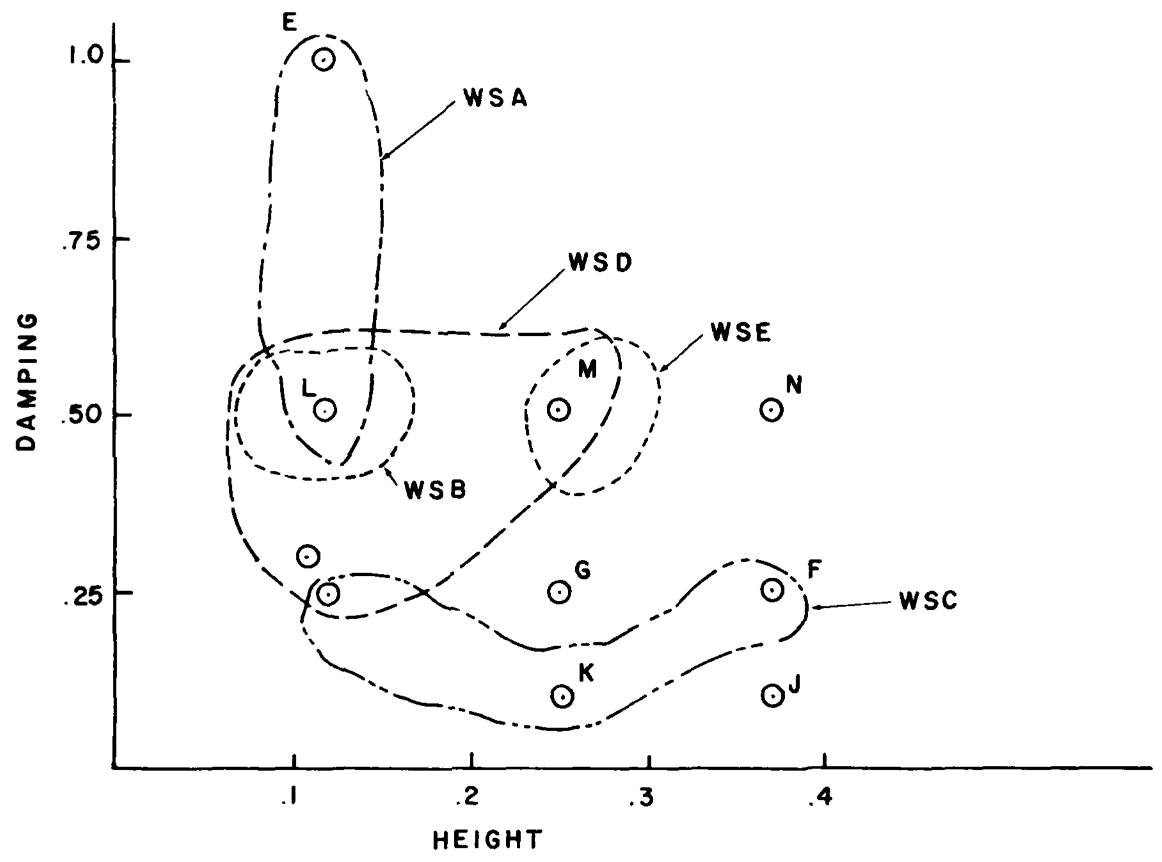


Figure 41. Schematic Plot of Windscreen Locations in Damping-Height Space Based on Judgments Associated with the Dynamic Bar Target

When the analysis of the physical attributes is compared with the discriminate analysis of ranks, it seems that the two-dimensional representation on the physical scale yields more discriminatory power. The simple ranking approach apparently does not fully represent the dimensions to which subjects respond. Unfortunately, it cannot be determined whether improving the sensitivity of the CGI representations will improve windscreen classification without further work in the CGI scales.

SECTION VI

SUMMARY AND GENERAL CONCLUSIONS

The studies reported herein utilized two basic psychophysical techniques, magnitude estimation and stimulus matching, to judge windscreen-induced distortion. Magnitude estimation was used in conjunction with many types of targets to identify those which best afforded good discrimination among windscreens with respect to distortion. The matching procedure was utilized in only one demonstration study to illustrate that an alternative psychophysical method could be used to assess distortion. The results of the studies are reported in the appropriate sections of this document along with observations and will not be repeated here. Rather, it is desirable to make several general comments and note some methodological observations in this final section of the report.

The Phase I studies that involved several subjects show that certain targets offer promise when the magnitude estimation technique is employed. Experiment I, which utilized six subjects, demonstrated that grid targets of the appropriate size are useful for scaling distortion. The work conducted on magnitude estimation suggests that any target having rectilinear or periodic line elements will be superior to aperiodic targets and to ones that do not have linear elements. It seems that human perceptual systems utilize linear patterns in establishing a set or "expectancy" about what should be perceived; deviations from this expectancy are enhanced, contributing to the perceived distortion. This is not surprising when one considers how visual perception functions. The human observer performs best as a "null" instrument. That is, differences from some neutral or standard values are more readily perceived than absolute values.

Dynamic targets can also be used to facilitate judgments of windscreen distortion. This approach works both for magnitude estimation (Experiment VIII) and for the CGI matching procedure. The dynamic target accentuates distortion by generating transformations of the optical array over time. It should be emphasized that extremely complex patterns may not be useful as dynamic targets. A horizontal bar grating and a single bar both function

well under certain circumstances, although there may be some indication that a single bar is to be preferred. In general, the dynamic mode is best with simple target forms. Too much movement in the visual field masks or obscures the distortion. For example, when moving checkerboard targets were screened, it was found that they induced undesirable perceptual effects that made judgments difficult. From a methodological point of view, the major deficiency in the dynamic approach is that it requires some use of memory in making judgments about distortion as only selected portions of the windscreens are individually and sequentially highlighted at any moment in time.

Two of the studies showed that a linear combination of judgments obtained with monocular and binocular viewing could produce good windscreen classification. It is not surprising that monocular and binocular viewing yield different patterns of discrimination. The nature of the windscreen distortions causes a difficult set of optical transformations to be presented to each eye. Thus, it seems likely that retinal disparity effects enable the observers to perceive aspects of the distortion that are different from those available in monocular viewing. This finding raises that issue of whether or not depth effects should be incorporated into the target display. Introducing depth effects into targets was deliberately avoided because it was felt that they might unnecessarily complicate the testing procedure. It is now apparent that were windscreen testing to be continued, it may be advantageous to investigate depth in the form of stereopsis. That is, a target that could be perceived only with stereoscopic fusion might reflect distortions better than simple two-dimensional arrays. Of course, it is difficult to know if depth effects are actually the phenomena involved in the canonical variates that result from monocular and binocular viewing. The analytical solution is merely a linear combination of variables that may not have any simple, meaningful physical interpretation.

The problem of subject variability has been mentioned often in this report. An analysis of the standard deviations of subjects that participated in several studies over a period of months showed that, in general, subjects did not alter or improve their discrimination across replicates or trials. For example, when five experienced subjects were rank ordered on the basis of their standard deviations for four different studies, only two minor

reversals of ranking were observed in two studies. Highly reliable subjects remained so and variable subjects did not improve. It is believed, therefore, that future studies of this type should focus on screening subjects rather carefully. For example, subjects could be tested on magnitude judgments for physical stimuli so that objective criteria of performance would exist. Having objective criteria would also facilitate training of subjects, if desirable. The problem which occurred, of course, was that no reliable physical criteria of distortion were available to be used in assessing the subjects' performance (see Appendix C).

Another issue intimately associated with the variability problem is sensitivity of subjects to the variable of interest. There was no way to assess subjects' absolute or difference thresholds for distortion. Since no known increments of distortion could be produced, there was no simple procedure whereby subjects could be tested to see if they possessed sufficient sensitivity to distortion to be useful in the task. Some distinctions among windscreen distortions are very subtle and require careful visual inspection to perceive differences. The precision of the work conducted might have been greater had some mechanism been available for choosing only those subjects who possessed good distortion "acuity". Alternatively it is possible that some increments of distortion were simply below subjects' detection thresholds. This would account for some of the variability and failure to discriminate among distortions.

The present work was severely limited by the lack of available knowledge about human perception of distortion. Certain physical measures of distortion do not appear to correspond with psychological judgments (see Appendix C)*. At present, there is no psychological scale of distortion that can be related to physical indices. This means that the suitability of the windscreen sampling for representation of distortion had to be assumed. Based on the windscreen classification data, it seems that the sample used did not best represent the full range of distortion. One windscreen contained so much distortion that it may have biased subjects' judgments and resulted in their

*This was true for these experiments. Gomer and Eggleston (1978), however, found that magnitude estimates of distortion did correlate well with displacement grade when a slightly different procedure was used. It is possible that other physical measures might show better correlations with our psychological judgments.

failure to effectively rate the others. That is, subjects may have used only a coarse, nonsensitive part of their psychological scale of distortion. The problem is analogous to making absolute judgments of weight when one sample weighs 10 pounds and the others weigh 2, 2.5, and 3 pounds. The similar samples tend to be classed together and discrimination is not as sensitive as it could be. Without some prior knowledge of how humans perceive distortion and information about the size of the increments of the psychological scale, there were no a priori grounds upon which to accept or reject any windscreen for the sample used. The sample had to be selected from what was available, on an intuitive basis.

The problem is further complicated because the evidence from our studies suggests that distortion is a multidimensional attribute. It is very unlikely that any single measure can capture the full global aspects of distortion in a magnitude estimation procedure. In fact, the evidence (Experiment VI) points to the need for multiple, independent assessments of distortion. This means that an observer would have to make judgments on windscreens using two or more targets. The judgments would then be weighted according to the canonical coefficients and evaluated by comparison with the canonical variates of acceptable windscreens. The procedure would require that each visual inspector be "calibrated" on a known set of windscreens in order for the comparison to be meaningful.

The CGI work suggests that there may be an alternative to the above procedure that would also reflect the multidimensional qualities of distortion. The matching techniques employed have the advantage that observers are used in their most sensitive mode--that of a "null" meter. The success of the procedure depends upon the CGI stimuli; they must represent approximately equal intervals on the subjects' psychophysical scale of distortion. Again, since this information was not available, the CGI stimuli had to be selected on an intuitive basis; hence it is likely that the precision of the technique can be increased with suitable development of a psychophysical scale of distortion.

It now seems clear that further development of techniques for assessment of distortion requires a basic knowledge of how observers scale distortion.

Without these basic psychophysical data, it will be difficult to improve upon any visual inspection technique. For this reason, we have not outlined a detailed visual inspection procedure. It would be possible to calibrate inspectors using the five test windscreens and apply the canonical weights for particular targets to use the present magnitude estimation data as a basis for inspectors' judgments on newly manufactured windscreens. Such an approach seems unwarranted at this stage of investigation, however, because a fundamental understanding of the perception of distortion does not exist at present.

Our work on magnitude scaling was task-focused. That is, we wanted to test different target configurations for their efficiency in highlighting distortion and then use the better targets as a vehicle to aid in quantifying windscreen quality. After the Phase II work, it became clear that the state of knowledge about human perception of distortion was insufficient for us to design an "optimal target". We therefore focused on the issue of scaling and developed the CGI matching procedure to demonstrate that development of a scale is possible. With further refinements, one could better understand the mechanisms underlying the perception of complicated distortions. It would then be feasible to undertake development of appropriate inspection techniques.

The work summarized in this report warrants the following conclusions:

(1) While several types of static and dynamic targets facilitate judgments of distortion; there is compelling evidence to suggest that rectilinear target types are the most appropriate for use in judging distortion. The grid targets currently employed in visual inspection appear quite suitable for the task.

(2) Depth effects in target displays may facilitate judgments of distortion and should be explored further.

(3) A screening and/or training procedure for inspectors should be developed.

(4) Development of the psychophysical scale(s) used in distortion judgments will have to <u>precede</u> improvement in visual inspection procedures of optically distorting media.

(5) The reliability of subjects' judgments can be highly variable and must be considered when data are interpreted. This variability may be reduced with appropriate training prior to the subject's use as an inspector of windscreen optical quality.

APPENDIX A

Instructions to Subjects for Static Targets:

You will be presented with several aircraft windscreens in a random order. Your task is to judge the summetry of the target pattern viewed through each windscreen. Sometimes the target will appear distorted in various places. Some of the patterns will not appear square; they will not be symmetrical. Your judgment should be made by first observing the standard (i.e., viewing the target without a windscreen installed) and comparing it to the test windscreen. The standard is completely symmetrical. Your task is to judge what decimal, fraction, or percentage of the test pattern appears symmetrical. If you think all are symmetrical, then you will report "all" or "100%" or "1.0". If you think only half are symmetrical, then report that and so forth. You may use any rating scale with which you feel most comfortable. Your rating scale, for example, may include decimals, fractions, or percentages. We do not want you to count the number of nonsymmetrical pattern elements. We are interested in a global impression. Your first impression is probably the best.

Do you have any questions?
(Answer questions)

OK, here is a photograph of sample pattern. Look at it and make your judgments. Here's another. Try this one.

(If S appears to understand task, then take him into experimental room and begin.)

Instructions to Subjects for Dynamic Targets:

As in previous studies, you will be viewing targets through aircraft windscreens. The targets in this study will be horizontal bars (square wave gratings) that move up and down. Your task will be to judge how much of the horizontal bar pattern distorts as the pattern moves up and down. This means that you should watch the pattern cycle up and down a few times and then make an overall judgment about how much of the total pattern is distorted.

Your overall general impression is probably your best judgment; you will do better if you do not try to count the bars or keep any precise mental records. We just want your overall impression of the total distortion the pattern undergoes as it cycles up and down.

Do you have any questions?

Instructions to Subjects for CGI Matching Study:

I. Ordering Set

Here is a set of computer generated images. We would like you to arrange them in an order that makes sense to you. Use any or all of the dimensions in the image that seem important. Naturally, we want you to focus your attention on the nature of the images, not on other factors such as the quality of the Xerox.

II. Windscreen Judging

We are going to show you several windscreens and ask that you carefully view the target (grid) through each one. After you have looked through all of them and have some idea of the range and kinds of distortions they produce, we will then randomly select one windscreen and ask you to make the following type of judgment: using the set of computer generated images you have just ordered, we want you to select the image that best seems to represent the windscreen through which you are viewing the target.

APPENDIX B

DATA TRANSFORMATIONS

In the general description of the analytical procedure at the beginning of this report, it was noted that the discriminant analysis procedure requires that errors be normally distributed. The distributions of the deviations from the means (errors) for subjects' judgments of several targets have been analyzed. Data from one of these, the static grating, are presented in Table 39. The skew and kurtosis* were calculated for three kinds of scoring: raw scores with no transformation, Z-scores only, and Z-scores of an arcsine square root transformation. Note that grating 3 shows a substantial improvement for the arcsine square root transformation. This was also the grating target that produced the best discrimination in that study.

TABLE 39. SKEW AND KURTOSIS FOR TRANSFORMATIONS OF THE DATA FROM THE STATIC GRATING STUDY

	Grating Condition	Skew	Kurtosis
No Transformation	1	.226	-.606
	3	.423	-.749
Z-Score Only	1	-.648	1.585
	3	-.091	-.135
Z-Score of Arcsine Square Root	1	-.553	.407
	3	-.015	-.378

These grating data are presented because, of the Phase I studies, they represent a somewhat unusual case. Of the three subjects in this study, one used a very small part of the proportion scale and concentrated all his judgments at the low end. Since this person was very consistent and indeed did discriminate, his transformed Z-scores contributed to, instead of confounding, the analysis of windscreen distortion. The skew in the distribution of errors for the grating study was largely removed by the selected transformation despite this subject's unusual response range. The arcsine square

*For a normal distribution, skew = 0 and kurtosis = 1.

root transformation performed equally well when data were skewed less dramatically.

As a check on the utility of the transformation, several discriminant analyses were run on Z-scores only. In general, discrimination was not as good as when the arcsine square root transformation was applied before analysis.

APPENDIX C

CORRELATIONS OF PHYSICAL AND SUBJECTIVE MEASURES

One of the purposes of the present work was to compare physical and psychological measures of distortion. For this analysis, two physical measures were obtained, lens factor and displacement grade, for each of the five windscreens studied. Each of these two measures was correlated with the Z-score arcsine square root judgments obtained for the discriminant analysis. Pearson produce moment correlation coefficients were calculated for selected targets and are reported in Table 40.

TABLE 40. PRODUCT-MOMENT CORRELATIONS FOR PHYSICAL AND PSYCHOLOGICAL MEASURES OF WINDSCREEN DISTORTION

	Physical Measure	
Target Type	Displacement Grade	Lens Factor
Grid (Exp. I)		
0.275° Black Lines	0.276	0.253
0.275° White Lines	0.134	0.075
Checkerboard (Exp. V)		
0.08°	0.179	0.186
0.37°	0.077	0.066
Photo Study, Judgments on Windscreens		
0.275° Grid, Black Lines	0.378	0.506
0.15° Horizontal Grating	0.290	0.440
0.37° Checkerboard	0.343	0.458
0.275° Grid, White Lines	0.390	0.497
Photo Study, Judgments on Photos		
0.275° Grid, Black Lines	0.073	0.168
0.15° Horizontal Grating	0.020	0.159
0.37° Checkerboard	0.260	0.270
0.275° Grid, White Lines	0.231	0.297

Overall, the correlations are rather low considering that the physical measures are used as one basis for acceptance or rejection of windscreens that come from the manufacturers. Lens factor tends to correlate most highly

with the psychophysical judgments; however, the highest correlation, 0.506, only accounts for about 25% of the variability in the data.

Since both the physical measures and the psychophysical judgments are supposedly measuring windscreen-induced distortion, it must be asked why the agreement is so poor. One possible explanation is that the magnitude estimation data is highly variable and consequently leads to low correlations. This is unlikely since the correlations are higher for studies that led to poor classification than they are for some of our better targets, such as the Experiment I grids. If the physical measures accurately reflect distortion, then correlations should be highest when physical measures are compared with the data from studies that produced the best windscreen discrimination. This was not the case. Therefore, it must be concluded that the physical measures do not accurately reflect the properties of distortion to which the human observers responded in this study.

APPENDIX D

RELIABILITY

Reliability is an index that indicates how well subjects tend to agree in their judgments of windscreen-induced distortion. When equal interval data are available, reliability can be calculated from the analysis of variance data. Formally, reliability* is defined as:

$$r_{xx} = 1 - \frac{s_e^2}{s_x^2}$$

where

s_e^2 = mean squares of the windscreen by subjects interaction

s_x^2 = mean squares for the windscreen effect

Table 41 shows the reliability for some of the targets tested.

TABLE 41. RELIABILITIES FOR JUDGMENTS ON SELECTED TARGETS

Target/Study	Reliability
Grids (Exp. I)	0.987
Bars, Linear Motion (Exp. XII)	0.986
Checkerboards (Exp. V, 3 subjects)	0.989
Checkerboards (Exp. XI, 10 subjects)	0.746
Photo Study (Judgments on Windscreens)	0.949
Photo Study (Judgments on Photos)	0.976
CGI Study	0.933

*Edwards (1962) also gives computing formulae for reliability when data are based on ranks.

It is interesting that most of these reliabilities are high even though windscreens were not accurately classified in some of the studies. This is probably a result of the large effect windscreen C had on the data; in effect, it inflated the mean squares for the windscreen effect. This explanation can be verified by eliminating windscreen C data from the analysis. This approach is not strictly correct because the remaining data do not necessarily reflect subjects' distortion judgments and hence yield meaningful reliabilities. That is, the biasing effect windscreen C exerts would still exist in the remaining data (see Section VI). Nevertheless, we have calculated the reliabilities when windscreen C is omitted from the analysis. These data are shown in Table 42.

TABLE 42. RELIABILITIES FOR JUDGMENTS ON SELECTED TARGETS WITH WINDSCREEN C OMITTED

Target/Study	Reliability
Grids (Exp. I)	0.967
Bars, Linear Motion (Exp. XII)	0.794
Checkerboards (Exp. V, 3 subjects)	0.955
Checkerboards (Exp. XI, 10 subjects)	≅0
Photo Study (Judgments on Windscreens)	0.914
Photo Study (Judgments on Photos)	0.567
CGI Study	0.811

It is apparent that windscreen C contributed substantially to some of the reliabilities shown in Table 41. Linear motion bars and the Phase II checkerboard study both yielded data that were strongly biased by the large distortion of windscreen C. While all reliabilities were lower when windscreen C was excluded from the analysis, most studies were largely unaffected. In the photo study, subjects were more reliable when they rated actual windscreens than when the ratings were made on photographs.

REFERENCES

Dixon, W. J., BMD Biomedical Computer Programs, Univ. of California Press, Berkeley, 1977.

Edwards, A. L., Statistical Methods for the Behavioral Sciences, Holt, Rinehart, and Winston, New York, 1962.

Gomer, F. E., and Eggleston, R. G., Perceived magnitudes of distortion, secondary imaging, and rainbowing in aircraft windshields. Human Factors, 1978, 20, 391-400.

Marriott, F. H. C., The Interpretation of Multiple Observations, Academic Press, New York, 1974.

Rao, C. R., Linear Statistical Inference and Its Applications, Wiley, New York, 1965.

*U.S.Government Printing Office: 1979 — 657-084/337